

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 20:10:33 ; Search time 38.03 Seconds

(without alignments)  
1274.701 Million cell updates/sec

Title: US-09-938-330-20

Perfect score: 6954

Sequence: 1 MAPRLALSYLLPLHCAICT.....PTGPTLCVLPDQSLRGHT 1252

Gapop 10.0 , Gapext 0.5

105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.1

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	3614	52.0	1201	AT53_HUMAN	O15072 homo sapien
2	3326	47.8	1205	AT52_BOVIN	P79331 b adams-2
3	3299.5	47.4	1211	AT52_HUMAN	O95450 h adams-2
4	1304	18.8	1077	AT10_HUMAN	O94324 homo sapien
5	1263.5	18.2	1629	AT59_HUMAN	O982n4 homo sapien
6	1244	17.9	1533	AT12_HUMAN	P58397 homo sapien
7	1119.5	16.1	967	AT51_RAT	O9wqg1 rattus norv
8	1108.5	15.9	968	AT51_MOUSE	P97857 mus musculu
9	1098.5	15.8	967	AT51_HUMAN	O9uh18 homo sapien
10	1068	15.4	930	AT55_HUMAN	O9una0 homo sapien
11	1062.5	15.3	997	AT57_HUMAN	O9ukp4 homo sapien
12	1040.5	15.0	860	AT56_HUMAN	O9ukp5 homo sapien
13	1027	14.8	930	AT55_MOUSE	O9u001 mus musculu
14	1013.5	14.6	890	AT58_HUMAN	O9up29 homo sapien
15	945	13.6	905	AT58_MOUSE	P57110 mus musculu
16	898.5	12.9	837	AT54_HUMAN	O75173 homo sapien
17	807	11.6	630	AT54_RAT	O9esp7 rattus norv
18	528	7.6	450	AT10_MOUSE	P58459 mus musculu
19	417.5	6.0	826	AD08_MOUSE	O05910 mus musculu
20	398.5	5.7	571	DISJ_BOVTA	P30431 bothrops ja
21	393	5.7	824	AD08_HUMAN	P78325 homo sapien
22	380.5	5.5	207	AT55_BOVIN	O9ct92 bos sapien
23	346	5.0	956	AD19_HUMAN	O9h013 homo sapien
24	341	4.9	920	AD19_MOUSE	O35674 mus musculu
25	328	4.7	774	AD28_MOUSE	O9j1n6 mus musculu
26	319	4.6	245	AT54_BOVIN	O9ct93 bos taurus
27	311.5	4.5	776	AD28_MACPA	O9x16 macaca fasc
28	309	4.4	788	AD07_MOUSE	O35227 mus musculu
29	306.5	4.4	776	AD21_MACPA	O28475 macaca fasc
30	306	4.4	722	AD21_HUMAN	O9uk18 homo sapien
31	302	4.3	790	AD30_HUMAN	O9uk12 homo sapien
32	297.5	4.3	760	AD25_MOUSE	O9rl59 mus musculu
33	295	4.2	1077	SM5A_MOUSE	O62217 mus musculu

## ALIGNMENTS

```

RESULT 1
ID AT53_HUMAN STANDARD: PRT: 1201 AA.
AC O15072;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 3) (ADAM-TS 3) (ADAM-TS3) (Fragment).
GN ADAMTS3 OR KIA0366.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
CC -1- DNA RES. 4:141-150(1997).
CC -1- COFACTOR: BINDS T-ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC -1- FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: AB002364; BAA20821.1; -
DR MEROPS: M12.220; -
DR MIM: 605011; -
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zn_MTPepidse.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF00090; TSP_1; 4.
DR SMART: SM00209; TSP1; 4.

```



RA MEDLINE=95348096; PubMed=7622483;  
 RA Collige A., Beschien A., Samyn B., Goebels Y., Van Beemen J.,  
 RA Nussgens B.V., Lapiere C.M.;  
 RT "Characterization and partial amino acid sequencing of a 107-kDa  
 RT procollagen I N-proteinase purified by affinity chromatography on  
 RT immobilized type XIV collagen."  
 RL J. Biol. Chem. 270:16724-16730(1995).  
 CC -I- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR  
 CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO  
 CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN  
 CC COLLAGEN BIOSYNTHESIS.  
 CC -I- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain  
 CC alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains  
 CC at Ala-1-Gln.  
 CC -I- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -I- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO  
 CC COLLAGEN TYPE XIV.  
 CC -I- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
 CC MATRIX (BY SIMILARITY).  
 CC -I- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL  
 CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS  
 CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS  
 CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.  
 CC -I- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
 CC -I- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -I- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY  
 CC SIMILARITY).  
 CC -I- DISEASE: DEFECTS IN ADAMTS ARE THE CAUSE OF DERMATOSPARAXIS, A  
 CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN  
 CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN  
 CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.  
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC -I- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
 CC -I- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X96389; CAA65253.1; .  
 DR InterPro: IPR001762; Disintegrin.  
 DR InterPro: IPR002870; Pep\_M12B\_propep.  
 DR InterPro: IPR000884; TSP1.  
 DR InterPro: IPR000130; Zn\_MTPeptase.  
 DR Pfam: PF01562; Pep\_M12B\_propep; 1.  
 DR Pfam: PF01421; Reptolysin; 1.  
 DR SMART: SM00090; TSP\_1; 4.  
 DR SMART: SM00209; TSP1; 4.  
 DR PROSITE: PS50215; ADAM\_MPRO; 1.  
 DR PROSITE: PS50092; TSP1; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; FALSE\_NEG.  
 DR PROSITE: PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 DR HydroLase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.  
 FT SIGNAL 1  
 FT PROPEP 28  
 FT CHAIN 253  
 FT METAL 254  
 FT METAL 402  
 FT ACT\_SITE 403  
 FT METAL 406  
 FT METAL 406  
 FT METAL 412  
 FT DOMAIN 412  
 FT DOMAIN 474  
 FT DOMAIN 555  
 FT DOMAIN 611  
 FT DOMAIN 716  
 FT DOMAIN 717  
 FT DOMAIN 845  
 FT DOMAIN 846  
 TSP TYPE-1 2.  
 TSP TYPE-1 2.

FT DOMAIN 906 968 TSP TYPE-1 3.  
 FT DOMAIN 969 1024 TSP TYPE-1 4.  
 FT SITE 685 687 CELL ATTACHMENT SITE (POTENTIAL).  
 FT DOMAIN 31 35 POLY-ALA.  
 FT DOMAIN 177 180 POLY-GLU.  
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 943 943 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 987 987 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1025 1025 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1092 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1205 AA; 133887 MW; 7B5B232A45320371 CRC64;  
 Query Match 47.8%; Score 3326; DB 1; Length 1205;  
 Best Local Similarity 52.1%; Pred. No. 9,6e-166;  
 Matches 641; Conservative 160; Mismatches 351; Indels 78; Gaps 20;  
 QY 7 LLSYLLPLHCALCTAAGSRTPELHLSGKLSDYGVYVPCSTDFRGRLSHVYSGPAASAG 66  
 DB 17 LILLLLPLPADARLAAADADPPGCGPGHAEKILAVPRITDQGRVSHVSA-ATAPAG 75  
 QY 67 SMVYDTPPLPRHSHLRVARSPLHPGGLTLPGRVGRSHSLVFNVVEGKELRLRPNRR 126  
 DB 76 ---VRRRAAPAOIPGLS-GGSEEDPGG-----RLFYNTVVGRODLRLRPNRR 121  
 QY 127 IYVPGSVVQMODFRLLPQPLRQECVYTGVTGM-PCAANAISGGLAGLITDSTDF 185  
 DB 122 LVAPGATVMOEGSGATRYEPLGLTCLVGDVAGIAESSVALSCGLAGLIMEEERF 181  
 QY 186 FLEPLERGOEKEA-SGRTHVYVRRRAVQOEVAEEDG-----DLHNEAFGIGDLPNLTGL 239  
 DB 182 FLEPLEKGLAAKEADQGRVHYVYHRTISR--PPLGGPQALDITGISNDLSISRLGV 239  
 QY 240 VDDQGLDTPRK-RRIAKRGSTSEVLLVYDSDVYRFHKEHNVYVLTLMNYDEIYHDE 298  
 DB 240 LEEVNSRRRRRRRAADDDNIEVLLGVDSVVOFHTETHEVXYLLTLMIYVEIYHDE 299  
 QY 299 SLGVHINILVYLINWVGRKOSTSLERGNPSRSLKYOVRMAHSGOORDPSHAHHNVF 358  
 DB 300 SLGAHINIVLAILLSGKMSLSLEIGNPSOSLENNCRNAVLDQKPDTHDEYHDAIF 359  
 QY 359 LTRDFFGS---GVAVTGMCHPLRSCALNHEDEGFSSAFVIAHETGHVLMEMHDGQNGC 415  
 DB 360 LTRDFFGSGCMQGYAPVTGMCHPVRSCTLNHEDEGFSSAFVIAHETGHVLMEMHDGQNGC 419  
 QY 416 ADETSLSGVMAPIVOAARHFRHRSKSTLSRYLPSYDCLLDPPRAMPORPELPGIN 475  
 DB 420 GDEVALGSIMAPLVQAAHFRHRSKSOQELSRHSLYDCLLRDPFHMPALPOLDPLGH 479  
 QY 476 YSMDBCRFEDESGYONCLAFRTPEPKQLMCSHPDMPYCKTKRGGPDLDTGCACAGKMC 535  
 DB 480 YSMNOCRFDEGLGVMCTARFTDPCKQLMCSHPDMPYCKTKRGGPDLDTGCACAGKMC 539  
 QY 536 FKGRHIMKSPQYTGQDGSMSATKRFSGCSRGCGGYRHSRSCNNPSPAYGGRPLGLGM 595  
 DB 540 FKGRHIMLTPD-LIKRQGMQAMSPFGSCSRGTGTGKFRTRCDNPHANGRTSGCLA 598  
 QY 596 FEYVQNSNEECPTGYEDFPAQOCARNSYVYHONAKRSWYVPPDDADKCELTQSAQT 655  
 DB 599 YDFQVNSNODPDALADLPFRFECQKQWMLYEHEDADQHHMLPHHRRAKECHLYCESKET 658  
 QY 656 GDVVPANQVYHDGTRGSTRDPYSVACARGECPVGCDEKEMSGKADKCGVCGSDNSHCT 715  
 DB 659 GEVVSARMYHWDGTRGSTRDPAFSLCYVGDRCRKVCGCGVIGSSKQEDKCGVCGSDNSHCKV 718  
 QY 716 VKGTGLGASQAGALKVQVOPAGARHIOJEALEKPSHRSVYVKNQVYGSTILMPKRG-EAT 774  
 DB 719 VKGTFRSPKKLGIYKFEIPAGARHLLIOEDADTJSHLVAKNLETKFTLNEENDVDPN 778





943 THKMPKACAGRRPARRPCLLRVCPMPKMRGLAMSGCSATCGEIGDIOARVVCRTNANSL 1007Y  
890 VHRGFCALSLKPKAIRACNAPCEGSDPVWVITGEMPEPCSGTTCGRGMQVRSVCRTQPLDHN 949Y  
863 VQRKJCHKRRPPIRRRCNHRGSDPQWVITGEMGACSRGSGKIGVOTRIQCLPLSLNG 942Y  
853 VQRKJCHKRRPPIRRRCNHRGSDPQWVITGEMGACSRGSGKIGVOTRIQCLPLSLNG 942Y  
890 VHRGFCALSLKPKAIRACNAPCEGSDPVWVITGEMPEPCSGTTCGRGMQVRSVCRTQPLDHN 949Y

DR PROSTATE: PS00092, ADAM\_12, TSP1, 2.  
DR PROSTATE: PS00142, ZINC\_PROTEASE, 1.  
DR PROSTATE: PS00427, DISINTEGRIN\_1, FALSE\_NEG.  
KW Hydrolase: Metalloprotease: zinc; Glycoprotein; Zymogen;  
Repeat: Extracellular matrix

Query Match 18.8%; Score 1304; DB 1; Length 1077;  
 Best Local Similarity 31.0%; Pred. No. 8.9e-61;  
 Matches 359; Conservative 144; Mismatches 484; Indels 172; Gaps 44;

NON\_TER 1 1  
 PROPER <1 207 BY SIMILARITY.  
 CHAIN 208 1077 ADAMTS-10.  
 METAL 366 366 ZINC (CATALYTIC) (BY SIMILARITY).  
 ACT SITE 367 367 BY SIMILARITY.  
 METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).  
 METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).  
 DOMAIN 434 520 DISINTEGRIN-LIKE.  
 DOMAIN 578 679 CYS-RICH.  
 DOMAIN 680 802 SPACER.  
 DOMAIN 521 577 TSP TYPE-1.1.  
 DOMAIN 799 860 TSP TYPE-1.2.  
 DOMAIN 922 976 TSP TYPE-1.3.  
 DOMAIN 981 1031 TSP TYPE-1.4.  
 CARBOHYD 64 64 TSP TYPE-1.5.  
 CARBOHYD 196 196 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 CARBOHYD 297 297 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 CARBOHYD 714 714 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 CARBOHYD 769 769 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 CARBOHYD 866 866 N-LINKED (GLCNAc. . .) (POTENTIAL).  
 SEQUENCE 1077 AA; 118072 MW; 391ADE18DCBFB587 CRC64;

Query Match 18.8%; Score 1304; DB 1; Length 1077;  
 Best Local Similarity 31.0%; Pred. No. 8.9e-61;  
 Matches 359; Conservative 144; Mismatches 484; Indels 172; Gaps 44;

1 RSPDEFLS-SLESEYELAFPTFVDHNGALLAF-----SPPP-PRROR-- 40  
 85 VARSPLHGGTLNMGGRGRHSLYVNVVFGKELHLRLPRRLVGVSSWEMEDRELF 144  
 41 -----GT-----GATAESRLFYKVASPSTHLLNTRSSRLAGHVSVEVTRGLAW 88  
 145 ROPLEOCVYTGVTGVMGGA-VAISNCDIAGLIRFDSDFLEPLE--RGOQEKAS 200  
 89 QRAARPHLYAGHIGQAGSSSHVAISTCGGLHGLIVADEDEYILEPLHGPKRSRPEES 148  
 201 GRTVVVREAVQDEMAEPDGLINEARGLGD--LPNLGLYGDOLG-DTEKRRHAP 256  
 149 G-PHYVYKRSLLRHPHLDTACGVDEKPMKRPWMLRTKLPPRAPRLGNETER--GGP 203  
 257 G-----TEVLLVYDSDVVRPHGKEHONYVLTLMIVDEYIDSELSGHINIAIV 309  
 204 GKRSVSRREYVEYLVDADKMAVAYHGRDYEOYLAIMIVAKLFDDSLGSTVINIIV 263  
 310 RLIMVGYROSLIERGNPSRSLDQVCRMAHSQGRDPSHAER-----HDHYVF 358  
 264 RLILLLEDDQPLEITH-HAGKSLDSFCW--OKSIVNHSGHGNAIPENGVAHDTAVL 318  
 359 LRDPF-----GPSGAPVYTGKCHLRSALNHDEGSSAPVIVAHNGHYLGMEHDG 410  
 319 IIRYDIICYKKNPCSTIGLAPRGKMERSCSVNEDIGLPOATTIAHETGHTFGAMHDG 378  
 411 OGNGC-ADETSISYMAFLVQAARFHFMSRCSKLELSRYLP--YCLLDLDPDPAPQ 467  
 379 VGNSCGARGQDPAKIMAAHITIKTNPFWSSCNRDYITSLFDGIGLCLNNRPRQDFVY 438  
 468 PRELPGLNSMDQCRFDGSGSYOTCLAFRRFPCKQKLMCHPNPFFCKTKKGPPLDTG 527  
 439 PTVAPQADADADOCRFHGVKSRQC--KYGEVCSSELMCLSKNR--CTNNSIPAEGT 493  
 528 EAPGK---WCFKGHCT--WKSPPQYGGDGMSSWTKFSGSCGSGGVRSRSCNN 581  
 494 LQDHTIDKGMCIKRYCVFPGSRPE---GVGDAMCPPTPMWDCRITCGGSSSSSRCDG 550  
 582 PSFAYGPRCLPMFEYGVYCNSECPGYEDEFRAQCAK-----RNSYVHQAQKSWV 635  
 551 PRTIGKCYCLGERRRHSNCNTDPCPSQDFREYQCESEFDSIPFRGFY-----KWK 603  
 636 PVEPDDAOKCELICOSADTGVYPMQVYHVDGTGCTGRDYYSVCAKGECPVAGCJAEV 695

604 TYR-GGGVACSLTSLAEFNFETERRAAVVDGTECR--PDTVVDICVSGECHVACDRILG 661  
 696 SMADKRCVCGGDNDHCTVGTIGKASKQALKLVQIPAGARHIQIEALESPPRSV 755  
 662 SDIREKCRVCGGDGACETIEGVSPASPGAGYEDVWIRKGSYHITQDNLNLSHLA 721  
 756 VKNQVGSFLLNFKGKEATSRFTAMGLEWEDAVDAKESLKTSGPLDEAIIALLPTE 815  
 722 LKQDQESLLEGLRPGTPQHRPLPLAGTTPQLRQPDQVQSLPALINASTIVWLARTE 781  
 816 GGRSSSLAYKY--VIHEDLLPLIGSNVLLIEMDTYEWALKSMAPCSACGGGIQFTKG 873  
 782 ---LPALRYFNAPRIARDLIP-----PYSWHTAPWTQCSAQCGAGGSOVAVE 825  
 874 CRRRRHHWQRLTCD-HKKRPPIRRRCNPPCSQPVVTEWEGACSRSGCLGVOTRG 932  
 826 CRNQLDSSAAPPYCSAHSKLPKR-ORACNTEPC--PRDWVGNMWSLGRSC-DAGVRSR 882  
 933 IOCLPLSLNCTHRYMPAKACAGDRPARPPCLRVPCPAQWRLGAMSQCSATCGEIQOR 992  
 883 VVCGRRVSAAEKALDSDACQPRPPVLEBACHGPTCCPPEMALDMSSECTPSCGPLRRHV 942  
 993 VVCTNANSL-----GRCF-GDRPDIVQVSL-----PAG-----CGGNHONSTV 1030  
 943 VLCKSADHRAITLPPAHCSPPAKPPATMRCNLRRCPPARVAVGEWEGSCAQCGVGRORSV 1002  
 1031 RADVWEIGTEPGQWVPO-SGPLHPINKISGMAE--PCTGD-----RSVFCOME 1077  
 1003 RCT-----SHTGQASHCTALRP--PTTQCCAKAKDSPTPGDGPBECKDVKAVCYPLV 1055  
 1078 VLDRYCSIPGYHRLCYSC 1096  
 1056 LKFOCSRAVFRWCMCKCTC 1074

RESULT 5  
 ID AT99\_HUMAN STANDARD; PRT; 1629 AA.  
 AC O9P2N4; O9NR29;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).  
 GN ADAMTS9 OR KIAA1312.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
 RC TISSUE-Fetal;  
 RX MEDLINE=20396138; PubMed=10936055;  
 RA Clark M.E., Kehler G.S., Turbeville L.A., Boyer A., Arden K.A., Maki R.A.;  
 RT "ADAMTS-9, a novel member of the ADAM-TS/Metalloprotein gene family";  
 RL Genomics 67:343-350(2000).  
 [2]  
 RP SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).  
 RC TISSUE-Brain;  
 RX MEDLINE=20181126; PubMed=10718198;  
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro";  
 RL DNA Res. 7:65-73(2000).  
 CC -1 CORFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -1 SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).  
 CC -1 ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.

CC - TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.  
 CC EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG,  
 CC PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN  
 CC COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR  
 CC THYMUS.  
 CC - DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY  
 CC SIMILARITY).  
 CC - PFM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY  
 CC SIMILARITY).  
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC - SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
 CC - SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF261918; AA89106.1; -  
 CC EMBL: AB037733; BAA92550.1; -  
 CC MIM: 605421; -  
 CC InterPro: IPR001762; Disintegrin.  
 CC InterPro: IPR002870; Pep\_M12B\_propep.  
 CC InterPro: IPR001590; Reptolysin.  
 CC InterPro: IPR000884; TSP1.  
 CC InterPro: IPR000130; Zn\_Mtpeptidase.  
 CC Pfam: PF01562; Pep\_M12B\_propep; 1.  
 CC Pfam: PF01421; Reptolysin; 1.  
 CC SMART: SM00209; TSP\_1; 12.  
 CC PROSITE: PSS00215; ADAM\_MEPRO; 1.  
 CC PROSITE: PSS00427; DISINTEGRIN\_1; FALSE\_NEGS.  
 CC PROSITE: PSS0092; TSP1; 9.  
 CC PROSITE: PSS0142; ZINC\_PROTEASE; 1.  
 CC PROSITE: PSS0142; ZINC\_PROTEASE; 1.  
 CC Hydrolase: Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 CC Repeat; Extracellular matrix; Alternative splicing.  
 CC SIGNAL  
 CC PROPEP 19 287  
 CC CHAIN 18  
 CC DOMAIN 288 1629  
 CC DOMAIN 509 587  
 CC DOMAIN 589 642  
 CC DOMAIN 645 752  
 CC DOMAIN 753 880  
 CC DOMAIN 999 1053  
 CC DOMAIN 1056 1108  
 CC DOMAIN 1111 1156  
 CC DOMAIN 1184 1239  
 CC DOMAIN 1240 1295  
 CC DOMAIN 1332 1383  
 CC DOMAIN 1386 1439  
 CC DOMAIN 1445 1498  
 CC DOMAIN 1501 1554  
 CC DOMAIN 1562 1612  
 CC DOMAIN 88 96  
 CC SITE 223 223  
 CC METAL 434 434  
 CC ACT\_SITE 435 435  
 CC METAL 438 438  
 CC METAL 444 444  
 CC CARBOHYD 112 112  
 CC CARBOHYD 135 135  
 CC CARBOHYD 271 271  
 CC CARBOHYD 749 749  
 CC CARBOHYD 840 840  
 CC CARBOHYD 1213 1213  
 CC CARBOHYD 1267 1267  
 CC CARBOHYD 1267 1267  
 CC VARSPPLIC 1064 1072  
 CC VARSPPLIC 1073 1629

FT CONFLICT 367 367 F -> L (IN REF. 1).  
 SQ SEQUENCE 1629 AA; 182649 MW; C1C4CEFF58B8941F CRC64.  
 Query Match 18.2%; Score 1263.5; DB 1: Length 1629;  
 Best Local Similarity 28.0%; Pred. No. 17e-58;  
 Matches 357; Conservative 171; Mismatches 452; Indels 297; Gaps 49;  
 21 AGSGRPPELH-----LSKSLDGVCTVPCSDPRGRGLSHVSGSPAASAGSNVDTPEP 75  
 26 AAAYKRLRHPROVKLETLDEIVSPIRYNAUG-----BEFPT 65  
 76 IPRHSHLRVARSPLHPGCTLMPGRVGRHS-----LYFNVTYGRKELRLRPNRLVY 129  
 66 -----VWHFKRTRRSINSATDWPAPAFSSSSSTSSQAHYLSAFQOFLFNLANAGFLA 121  
 130 PGSSVE-----WQDFPELRPRQPCVYTGVTGPGAAVAISNCGLAGL 177  
 122 PLEVTLLGTGVNQTFRYSEEAEL-----KRCFYKGVVNTNSEHTAVISLCSGLGT 175  
 178 IRTDSTDFEIEPLER--GQDEKASGRTHVYREAVQOEMAPDDGLH----- 224  
 176 FRSHDGVFIPELPSMDEQDEEBQNKPHILYRSAPQ---EPSTGRACDTSEKRNH 232  
 225 -----NEAFGLGDLPLLGLV-----GDQLGDTERRRRRAKPGSY----- 259  
 233 SKDKKTRARKGGRINAGDVAALNSGLATEAFSAVGNKTDTRKTRTRKRLFLSYE 292  
 260 -SIEVLVVDSDVYRFPGKEHYONYVLTLMNIDEIYHDSLSGVHNLAVRLIMGYQ 318  
 293 RFVEVLVADNRKMYTHG-ENLDYHTLLTMSIVASTKQDSISGLNINIVMLIVIHNEQ 351  
 319 ---SLSLIERGNPSRLSEYOCRWMAHSOORQDPSHAHHHDHVVLTRODF-----GPSG 368  
 352 DGPSISF-----NAQTILKNFCOMQHS--KNSPG-GIHNDPAVLTRODICRAHDKDTLG 404  
 369 YAPYGMCHPLRSCALNHEDGSSAFVIAHETGHVLCMEHOGGNCADP--TSLGSVA 426  
 405 LAELGTTICDPRKSSISDESLTAFTIAHEIGHVNMPPD--DNCKCKEYGVSRQHWVA 463  
 427 PLVGAARFHRHMSKSKLESLRYLPS--YDCLDDPFPDAMPORPLPINSMDQCRF 484  
 464 PLTFEYINPMWMSKSKRYITFEFLDTGYGCLLNEPESRPYLPQVLPILYVVKQCEL 523  
 485 DFSGYGTCLAFTEPBCQKLMCSHPDNEY-FCKTKKGRPLDGTETCAPCKWCFKHCIMK 543  
 524 IFGGSGVVC---PYMOCRLMCMNNVNGYKKGCRTOHTPMADGTETCEBPCKHCKYGCVPK 580  
 544 SPEQTYGDDGWSMTRKFGSCRSRGGVRSRSCNNSPAYGRCPLGPMFEYQVONS 603  
 581 EMDYVP-TDSSWGSMSFPGTSCRTGGGKTAIRBCNRPENNGKRYCVGRMRKKSQNT 639  
 604 EECPGTYEDFRAOOCARKNSYVYHONAKH-----SNVP-YEPDDDAQKCELIQS 652  
 640 EPCIKQRDRDRBOCA-----HFDGKHFNINGLLPVNRWPKXSGILMDRCLPEYR 692  
 653 ADTGDDVFNMYVHDTRGSRYPYVSCARGECVPYGCCKEYGVSKADKCGVCGGDSH 712  
 693 AGNTAYIQLDRYIDTPGQ-QDTNDICVQGLCROAGCHVYNSKARDKCGVCGGDSNS 751  
 713 CRTVKGTLGKASKOAGALKVQIPAGARHIOLE---ALEKSPHRSVYKNQVTSFLNP 768  
 752 CKTVAGFEN--TVHYGVNVVRIPAGATIDVROHSFSETDDNYTLSSKGEFLNG 809  
 769 KGEKATSRFTTANG---LMEDAVEDAKESLTSGLPEPAIATIALPTEGGPSSLAAY 825  
 810 NFVVTAKKEIRIGNAVVEYSGS-ETAAVERINSTDRIOELLQVLL--SVGKLYNPVDR 865  
 826 YVIEHEDLLPLISNNVLLDEM-DYEW-ALKSNACSNACGGGIDFTYGGRRRRDHMY 883  
 866 Y-----SFTPIEDKPOQFYWMNSHGFWQACSPQGE-RKRLVCTRESIDLTV 913  
 884 QRHLCDHKRRPRPIRRR----- 900

```

Cc 914 SDORCDRLPQGHITTEPCGTIDCLRWNVASRSECSAOCGLGYFTLIDYAKTSRLDGKTE 973
Cc 901 -----CNOHP-----CSOPVWYEWGASRSCGKLGVOYTRIGIOLLPLSNGT 943
Cc 974 KYDDGFCSSHPKRSNKEKSGECNTGWRWSAITECSKSDG- GTRRRRAICV-----NTR 1028
Cc 944 HXWPAKACAGDBREARPCPLRVCPAOWRLGAMSCSATCGEGIOQROYVCFRANSLG 1003
Cc 1029 NDVLDSKCTKHOREKVTIQRCESEPCP-QMKSGDMSCLYTCGKHNRHROWECQFGEIDLIN 1087
Cc 1004 H--CEGD-REDTVOYSLPACGNGHONSTVRADWELGPEPGQWVQSGPLHPINKISSM 1060
Cc 1088 DRMCDDPETKFTSMOTCCQPCAS-----WQAG-FWGCQSVYTCGGGYL----- 1129
Cc 1061 CAEPCTGDSRVC-----QMEVLD-----RCSIDGYHRLCCVSCIKRASGP 1103
Cc 1130 -----FAVKCIIGTVMSVVDNDNCNATRPDTODCLPSCH----- 1166
Cc 1104 NPGPDPEPTSLPPSTP 1120
Cc 1167 -PPPAETRRSTYSAP 1182

RESULT 6
AFL2_HUMAN STANDARD; PRT; 1593 AA.
AC P58397;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-12 precursor (PC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=21264577; PubMed=11279086;
RA Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.;
RT "Identification, characterization, and intracellular processing of
RT ADAMTS-12, a novel human disintegrin with a complex structural
RT organization involving multiple thrombospondin-1 repeats.";
RL J. Biol. Chem. 276:17932-17940(2001).
CC -1 COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1 SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -1 TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -1 DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1 PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -1 PM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1 SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1 SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC -1 SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/

```

```

Cc or send an email to license@slb.ch).
Cc -----
Cc EMBL: AJ250725; CAC20419.1; -.
Cc DR MIM: 606184; -.
Cc DR PROSITE: PS02015; ADAM_MEPRO; 1.
Cc DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
Cc DR PROSITE: PS00092; TSP1; 2.
Cc DR PROSITE: PS00142; ZINC_PROTEASE; 1.
Cc KM Hydrolase; Metalloproteinase; zinc; signal; glycoprotein; zymogen;
Cc KM Repeat; Extracellular matrix.
Cc FT SIGNAL 1 25
Cc FT PROPEP 26 240
Cc FT CHAIN 241 1593
Cc FT DOMAIN 465 544
Cc FT DOMAIN 545 596
Cc FT DOMAIN 597 700
Cc FT DOMAIN 701 826
Cc FT DOMAIN 827 881
Cc FT DOMAIN 886 943
Cc FT DOMAIN 947 995
Cc FT DOMAIN 996 1315
Cc FT DOMAIN 1316 1364
Cc FT DOMAIN 1367 1423
Cc FT DOMAIN 1426 1471
Cc FT DOMAIN 1471 302
Cc FT SITE 302 305
Cc FT SITE 308 208
Cc FT METAL 392 392
Cc FT ACT_SITE 393 393
Cc FT METAL 396 396
Cc FT METAL 402 402
Cc FT CARBOHYD 105 105
Cc FT CARBOHYD 125 125
Cc FT CARBOHYD 215 215
Cc FT CARBOHYD 485 485
Cc FT CARBOHYD 685 685
Cc FT CARBOHYD 790 790
Cc FT CARBOHYD 951 951
Cc FT CARBOHYD 1104 1104
Cc FT CARBOHYD 1275 1275
Cc FT CARBOHYD 1300 1300
Cc FT CARBOHYD 1320 1320
Cc FT CARBOHYD 1371 1371
Cc FT CARBOHYD 1378 1378
Cc FT CARBOHYD 1503 1503
Cc SQ SEQUENCE 1593 AA; 177545 MW; 07F9F48B63BD83A3 CRC64;

Query Match 17.9%; Score 1244; DB 1; Length 1593;
Best Local Similarity 28.9%; Pred. No. 1.7e-57;
Matches 383; Conservative 185; Mismatches 554; Indels 202; Gaps 56;

Cc 30 HLSGLSDYGVVPCSTDRGRFLSHVVGSPAASAGSNVVDTPPLPHSHSLVARS 89
Cc 44 HFIKGLPEYHVVPVRVDASGHFLSYGLHYPTSSRRKRDLD----- 85
Cc 90 LHHGGLMPERVRNRHSLYEVVTFGKEHLRLPRNLRLVPGSSVWQEDFRELPR 149
Cc 86 ---GSEDW-----VYRISHEEKDLFNLTVNGFL--SNSYIMKRYGNLSHKM 132
Cc 150 QE-----CVYTGCV--TGMPGAAYAINSCDGLAFLPTDSTDFEIELEGRQOEASGRT 203
Cc 133 ASSAPLCHLSGLYLOQGTGRVTAALSAGHGLTFQLPHGDFIEPVKK-HPLVEGYP 191
Cc 204 HXVYRRAVOQEAEPDGLAHNAFGLGDLPLNLGLVGQDGLGTEKKR--RAAKPG----- 257
Cc 192 HXVYRQKV-PEKPE-----TCGLKDSVNT-----SOKOLWEKERNRLPERSIS 238
Cc 258 --SVS-----TEVLLYVDSDSVYRFGKEHGVONYVLTLMNIYDELYHDSLGVINIALVRL 311
Cc 239 RRSISKRWETLVADTKITEYHGSFNVSYILITIMNVTGFLNHPISGNALHIYVRL 298
Cc 312 IMV-GYROSILIERGNPSRSLBOVCMAHSGQRODPDSHAHHDHVYFLTRDQFGS--- 367

```

```

Db 299 ILLEEEQGLKIVH--HAERTLSSFCCKMOKSINPKSDLNPRPHHVAVLITRKIDOCAGENR 356
QY 368 -----GYAVYGMCIPIELNSCALNHEDGSSAFVIAHETGHVIGMEHDGQNGCADERSLG 422
Db 357 PETTLISHSGMCPHRSKININENSGPLAFTIAHIEHSGFIQHODKENDCEPYGRHR 416
QY 423 SWAPLVQAARHFRHMSRCKSLERLYPSVD-----CLDDPDPAMPORPELPGINYS 477
Db 417 YMSRLOVDPTPLTWKSCSEYITRFL---DRGMFCLDIDPKKKGLSKSVIAGVIYD 473
QY 478 MDGCRDFGSGYQTCLAFRTFEPCKOLMCSHPDNPYFCCKTKRKPILDGTCEACAGKCKCF 537
Db 474 VHHOCOLGYPNATFCQEVENV--CQTLACSVKG---FCRSKLDAAADGTGCGKKKMA 528
QY 538 GHCI--WKSPEOTYQODGMSWTKFSGCSGSGGVASRSRSCNNDSPAYGRRPLCGPM 595
Db 529 GACITVKKPEST---PGMGWMSHWSHSCSRCTGAGVQSAERLCNNPEPKFGKGYCTGER 585
QY 596 FEYOVNSEECPTEDFRADQCAKRSYVYHONAKHSWV--YPPDDAOACELICQSD 654
Db 586 KRYRLCNVHPCSEAPTRQMGCSSEFD--VYKNEIYHMFITNP---AHPCELYCVRPD 641
QY 655 TGDVVFMMVYVADGTRC--SYDPTYSVARGCVYVGCDEKXVSKKADKCGVCGGDSHC 713
Db 642 GQFSKMLDAYLDGTPCEEGNSRWVCINGICKMWGCDYFIDSNATEDRCGVCLGDCSSC 701
QY 714 RTVKGTILKASQAGALVQIPAGARHIOEALKSPHRSVYVNO-----VTGSITLN 767
Db 702 QVVR--KMKRQKRGSGYVIGILPGARDIRWELIEGAGNELIASEDPERYUINGEFTD 760
QY 768 PKGKATSRRTTAMGLEWEDAVEDEAKESLKTGSPLEPAIAI--LALPTEGGPSSSLAYKY 826
Db 761 WNG-----NYKLGTAVYQYDRKGDLEKIMATPTNSVMIQLEFQVLTNG-----IKXY 810
QY 827 VTHEDLPLIGSNVYLEEMTYEMALKSNAPSGACGGIOFTKGCRRRDHMYOHR 886
Db 811 TIQKGL---DNDV--EOM--YFMQYGHMTECSVTGCTGIRQTAHC--IKKRGWAKAI 861
QY 887 LCDHKRKP1RRRCNOHPCSPQVWTEWAGACSRSCGKIGVOTRGICLLPISNCTHY 946
Db 862 FCDPEPTONGRQKCKHEKAC--PRWMAWGEWACATGPRGCKRYVLCIQTWVSDQAL 920
QY 947 MPARACAGDRREARPCLR--VPCPAQWRILCAMSQCSATVCEG10QRYVCRITVANSIGHC 1005
Db 921 PPTDCQHLKPKTLISCRNIDILCPSDMTYGNMSECSVSCGGVIRISVTCANKHDE--PC 978
QY 1006 EGDR--PTVQVCSLPACG-----GNHONSTVADVWELCTEGCQVWPQSGPL 1051
Db 979 DVTKRPNRSLCGLOCPSSRRVILKPKKGTISGKNPPTLAKVPVPTSRPRMLTPTGPE 1038
QY 1052 HPINKISSMCAEPCT---GD-----RSVFCOMEYLDRYCSIPGYHR---LCVSCJIK 1098
Db 1039 SMSSTFSAISSPPTASKKCDLGKQWQSSQTPPELSSRYLSTSTSQPILTSLSLI 1098
QY 1099 KASGPN--PGDPPGTS---LPPESTGSPPLPGQUDADAABEPGKPT---GSEDHQH-- 1148
Db 1099 QPSEENWSSSDGTGTSGLVATTTSGSGSLSSSRNPITPVPTFYVNTLTKGEMETHSGS 1158
QY 1149 GRATULPGALDTSPP-----GTQHPFAPE--TP--JPCASW--SIGSTTPG 1188
Db 1159 GEEREGEDKEDSNPYWTKIRVPGNDAPVESYHMLAPLPLPDLSSRESWMPFSTVMEG 1218
QY 1189 GLPMGWQTTPVPEDEKQGGEDLHRPG--TSLPAD--LP---GRPEPCHTGTFTTLCV 1241
Db 1219 LLP---SQRTP---TSETGP---RVEGMVTEKPAWTLPLGLGDHPSPSGKTANRNHLK 1269
QY 1242 LPRD 1245
Db 1270 LPNN 1273

```

```

ATSL_RAT
ID ATSL_RAT STANDARD: PRT: 967 AA.
AC 09WQ01; 09ER11;
AD 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 Precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA Little S.P.;
RT "Induction of a disintegrin and metalloproteinase with the
RT thrombospondin type I motif (ADAMTS).";
RT Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 18-967 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=20304099; PubMed=10847486;
RA Diamantis I., Luehth M., Hoesli M., Reichen J.;
RT "Cloning of the rat ADAMTS-1 gene and its down regulation in
RT endothelial cells in cirrhotic rats.";
RL Liver 20:165-172(2000).
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-1-LEU-1684
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COPACITOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -1- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
CC CIRRHOTIC LIVER.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FORIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF149118; AAD34012.1; -
DR EMBL; AF304446; AAG29823.1; -
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; ZN_M12Bpeptidse.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; TSP_1; 3.
DR SMART; SM00209; TSP1; 3.
DR PROSITE; PS50215; ADAM_MPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS50092; TSP1; 2.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.

```

Query Match 16.1%; Score 119.5; DB 1; Length 967;  
Best Local Similarity 29.2%; Pred. No. 3e-51;  
Matches 320; Conservative 151; Mismatches 393; Indels 231; Gaps 47.

QY	DB	PCSN	YWT	SP	LN	GH	BE	CL	MD	KP	QNP	I	K	PS	D	L	BP	GL	YD	AN	RP	CO	FT	FE	BS	TH	PC	-D	501																													
498	QY	TEFPCQ	LM	CS	HP	ND	NY	-	CT	K	K	G	P	P	D	G	T	E	C	A	F	K	K	C	F	G	H	I	M	S	P	D	T	Y	Q	--	DG	554																				
502	DB	AAS	T	C	S	T	L	M	T	G	T	S	G	L	L	V	C	O	T	K	H	F	P	A	G	T	S	C	G	K	A	C	V	N	T	D	M	N	H	E	T	P	H	G	S	561												
555	QY	W	S	S	T	K	T	G	S	C	S	R	C	G	G	V	R	S	R	S	C	N	N	P	P	A	V	G	C	P	C	J	G	M	F	E	Y	O	V	S	E	S	C	P	-	G	T	Y	E	F	613							
562	DB	W	C	P	M	G	P	G	D	C	S	R	C	G	G	V	Q	Y	T	H	E	C	D	N	P	V	P	K	K	G	A	C	K	V	R	R	S	C	I	E	D	C	P	D	N	N	K	T	F	621								
614	QY	RA	O	C	K	A	R	N	S	Y	V	H	O	N	A	K	H	S	-----	W	V	-	Y	E	P	D	D	A	O	K	C	E	L	I	C	S	A	D	T	G	V	E	N	O	V	664												
622	DB	R	E	O	C	E	A	N	H	E	F	-----	S	K	A	S	F	G	N	E	P	T	W	E	N	T	P	K	Y	A	G	V	S	P	K	R	C	L	T	C	E	A	K	G	I	G	F	F	V	L	O	P	K	676				
665	QY	V	H	O	T	R	C	S	T	A	D	E	P	S	V	C	A	R	G	E	V	C	P	G	C	D	K	F	E	V	G	S	M	K	A	D	K	C	V	O	G	G	D	N	S	H	C	R	T	V	K	T	G	L	K	A	S	724
677	DB	V	V	D	T	S	T	P	C	S	-	P	D	S	T	S	V	C	V	O	G	C	A	C	A	G	D	R	I	I	D	K	K	F	E	D	K	F	D	K	O	V	G	G	N	G	S	T	C	T	A	S	G	T	-	T	S	733
725	QY	K	O	A	G	A	L	K	V	O	I	P	A	G	A	R	H	I	O	I	E	A	L	E	S	K	P	H	R	S	V	V	K	N	O	V	T	S	F	-	I	N	P	K	K	E	A	T	S	T	F	I	A	M	G	782		
734	DB	T	R	P	C	H	D	I	V	I	P	A	G	A	T	I	E	V	K	-----	H	N	P	G	S	R	N	S	F	L	A	I	R	A	A	D	T	Y	I	L	N	E	T	L	S	T	787											
783	QY	L	E	M	E	-----	D	A	V	E	D	K	E	S	K	T	S	G	P	L	A	I	A	L	P	T	E	G	S	R	S	L	A	Y	K	V	I	E	D	L	832																	
788	DB	L	E	O	D	L	T	T	K	G	V	L	K	R	S	G	S	A	L	E	R	I	R	S	F	L	K	E	P	T	I	O	V	-----	826																							
833	QY	L	P	L	I	G	S	N	N	V	L	E	E	M	D	Y	E	M	A	L	K	S	M	A	D	C	S	K	A	C	G	G	I	O	F	T	K	Y	C	R	R	R	D	H	M	O	R	L	C	D	H	K	892					
837	DB	-	M	T	G	-----	H	A	L	R	P	K	I	K	Y	T	F	-----	844																																							
893	QY	R	P	K	I	R	R	C	N	O	H	P	C	S	O	P	W	T	E	M	G	A	C	S	R	S	C	K	L	G	V	O	T	G	I	O	C	L	P	L	S	N	G	H	K	W	P	A	K	A	S	952						
845	DB	K	T	E	P	-----	N	A																																																		





Db 630 HNEF-----SKASFGNEPTVEMTEPKYAGVSPKDRCKLTCEAKGIGYFFVLQPKVVDCTPC 684  
 QY 672 SYRDPYSVCAGCEVPCDCEKVGSMKADRCYGGGDNHCHTIVKGLTGAKSKOAGALK 731  
 Db 665 S-PDSTSVCGOQCKYKACDRIIDSKKFFDCGCGGSGCKMAGSYV--TSRPGYHD 741  
 QY 732 LVQIPAGARHIQIEALEKSPHRSVYKNOVTSF--ILNPKCKEATSRFTFMGLMEF--- 786  
 Db 742 IYTIIPAGATNIEVK-----HRNORGSRNNGSFLAIRAADSTYILINGNFTLSTEDQDLY 795  
 QY 767 -----DAVEDAKESLTSQGLPEAPAIILALPTEGGPRSSLAKEYIHEDLLPLIGSN 839  
 Db 796 KGTVLRYGSSSAALERTISFSPLEPLITQVL-----MG-- 830  
 QY 840 NVLLEEMDTYEWALKSMAPCSKACGGGIQFTKYGCRRRRHHMVORHLCHKKRKPPIRR 899  
 Db 831 -----HALRPKIKFTYF-----KKKTESF-- 850  
 QY 900 RGNQHPCSOPVAVTEEMGACSCSKGLGVQRTGICLLPLSNGTHKVPAPAKACGD-RPE 958  
 Db 959 ABRPCLRVCPAPOMRLGAWSCSATCGEGIQGRQYVCTNANSLGHCEDRP-----DTV 1013  
 QY 1014 QVCSIPAC 1021  
 Db 901 STRPCADLFCP-HMVGDMSPSCSKTCGKGYKRTIKVSHDGVLSNESCDPLKPKHYI 959  
 QY 960 DECTLTQC 967

RESULT 9  
 ATSL\_HUMAN  
 ID ATSL\_HUMAN STANDARD; PRT: 967 AA.  
 AC Q9UHB8; Q9UHB8; Q9P2K0; Q9NSJ8;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).  
 GN ADAMTS1 OR METH1 OR KIAA1346.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Casas C., Pritchard M.A., Estivill X., Arbones M.L.;  
 RT "Cloning, characterization and mapping on human chromosome 21 of the orthologue of murine Adamts-1.";  
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RP SEQUENCE FROM N.A., AND FUNCTION.  
 RA TISSUE=Heart;  
 RA MEDLINE=99367466; PubMed=10438512;  
 RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S., Lombardo M., Iruela-Arispe M.L.;  
 RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity.";  
 RL J. Biol. Chem. 274:23349-23357(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Endothelial cells;  
 RX MEDLINE=20247184; PubMed=10785405;  
 RA Glenske J., Schmitt A.O., Pilarsky C., Hinzmann B., Weiss B., Rosenthal A., Thierach K.H.;  
 RT "Differential gene expression by endothelial cells in distinct angiogenic states.";  
 RL Eur. J. Biochem. 267:2820-2830(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20181126; PubMed=10718198;

RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro.";  
 RL DNA Res. 7:65-73(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=20289799; PubMed=10830953;  
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Tada T., Choi D.-K., Soeda E., Ohtsuka M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S., Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E., Minoshima S., Shimizu N., Nordstok A., Reichelt J., Kauer G., Bloeker H., Schaefer M., Schoen O., Desario A., Hennig S., Resselmann L., Dagand E., Ramser J., Beck A., Klages S., Hennig S., Resselmann L., Dagand E., Wehrmeyer S., Bozzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yaspo M.-L.;  
 RT "The DNA sequence of human chromosome 21.";  
 RL Nature 405:311-319(2000).  
 RN [6]  
 RP SEQUENCE OF 418-967 FROM N.A.  
 RC TISSUE=Melanoma;  
 RA Blum H., Bauersachs S., Mewes H.-W., Well B., Wiemann S.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CLEAVES AGGREGAN (A SIMILARITY). HAS ANGIOGENIC INHIBITOR INVOLVED IN ITS TURNOVER (BY SIMILARITY).  
 CC ACTIVITY. ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.  
 CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1938-GLU-1-LEU-1939 SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.  
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).  
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF170084; AAF15317.1; -  
 DR EMBL: AF060152; AAD48080.1; ALT\_INIT.  
 DR EMBL: AF207664; AAF23772.1; -  
 DR EMBL: AB037767; BAA92584.1; ALT\_INIT.  
 DR EMBL: AP001697; BAA95502.1; -  
 DR EMBL: AL162080; CAB82413.1; -  
 DR MIM: 605174; -  
 DR InterPro: IPR001762; Disintegrin.  
 DR InterPro: IPR002870; Pep\_M12B\_propep.  
 DR InterPro: IPR001590; Reptolysin.  
 DR InterPro: IPR000884; TSP1.  
 DR InterPro: IPR000130; Zn\_MTPeptide.  
 DR Pfam: PF01562; Pep\_M12B\_propep; 1.  
 DR Pfam: PF01421; Reptolysin; 1.  
 DR Pfam: PF00090; TSP1; 3.  
 DR SMART: SM00209; TSP1; 3.  
 DR PROSITE: PS50215; ADAM\_MEPRO; 1.  
 DR PROSITE: PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 DR PROSITE: PS50092; TSP1; 2.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; 1.

KW Hydrolyase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix; Heparin-binding.  
 FT SIGNAL 1 49 POTENTIAL.  
 FT PROPEP 252 50 BY SIMILARITY.  
 FT CHAIN 253 967 ADAMTS-1.  
 FT SITE 198 198 CYSTEINE SWITCH (POTENTIAL).  
 FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 402 402 BY SIMILARITY.  
 FT METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DOMAIN 476 559 DISINTEGRIN-LIKE.  
 FT DOMAIN 560 616 TSP TYPE-1 1.  
 FT DOMAIN 617 724 CYS-RICH.  
 FT DOMAIN 725 849 SPACER.  
 FT DOMAIN 850 908 TSP TYPE-1 2.  
 FT DOMAIN 909 967 TSP TYPE-1 3.  
 FT DOMAIN 843 846 POLY-LYS.  
 FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 227 227 P -> A (IN REF. 4 AND 5).  
 FT CONFLICT 468 468 O -> H (IN REF. 1).  
 FT CONFLICT 561 561 S -> N (IN REF. 1).  
 SO SEQUENCE 967 AA; 105383 MM; C189389324741BD1 CRC64;

Query Match 15.8%; Score 1098.5; DB 1; Length 967;  
 Best Local Similarity 29.2%; Pred. No. 3.8e-50;  
 Matches 300; Conservative 145; Mismatches 361; Indels 223; Gaps 46;

QY 87 RSPRLPGTLMGCRVGRSLFYNTVFGKELRLRPNRLVGVSSWEQEDFRELFRQ 146  
 DB 67 RAPGH--GTT--RLRLHA-----FDQDLLELRDPSFLAGFLQ--NVGKKSSST 113  
 QY 147 PLRQ-----ECYVTGGVTMGAAVAISMCGLITDSTDFTEP----- 189  
 DB 114 PLPELDLHACHTYSGVNDPSSAALSLCEBVGAFYLGEAFYTOPLPASERLATAP 173  
 QY 190 -----LENGQO-----EKEASGRTHVYRREAVQOE-----NAEPD 220  
 DB 174 GEKPPAPLQFLHRLRRNRGQDVGTCGVVDEPRPTKAEFDEDEGEDEGEDEGEPOD 233  
 QY 221 GDLNHEAFGLDLPNLGLVGDQIG--DTEKKRRKAKRGSTIEVLYVDDSVVRFKGEH 279  
 DB 224 -----PALOG--VGQPTGTGSIKKRFVSSHRY--VETMLVAVDQSMAFHG--SG 277  
 QY 280 VQNVVLTLMNIVDEIYHDESLGVHINIALVLLVNGVRSGLIERGNPERSLEYQCMA 339  
 DB 278 LKHYLTLEFVAARLYKHPISIRNSVSLVVLKLVHDEQKGPBY--ISNALTLIRNFCNM-- 335  
 QY 340 HSGQRDPS--HAHHNHVFLTRQDPS-----GYAPVTGMCHPLRSCALNHEDGFS 391  
 DB 336 -OKOHNPSSDRDAEHYDTALFLTRDQCGSQGSDTLGLMADVGVCDPSPRSQVIEDDQ 394  
 QY 392 SAFVIAHETGVLGMEHDGCGNCADETSLGS--VWAPLYQAARFRFMSRCKLELSR 448  
 DB 395 AAFVTAHELGVFMFMRD--DAKOCASLNGVNOSSHMAKSLSLDQSPWSPSAVMYTS 453  
 QY 449 YLPS--YDCLLDPEFDPAMPQPELPGLINSMDQCRFGSGSYQTCLAFRTFEPCKOLM 506  
 DB 454 FLNDSHGECLMDKPFQNI--QLPGDLPTGYDANRQCFHGEDSKNCP--DAASTCSTLM 510  
 QY 507 CSHNDNYF--CKTKKGPRLDTECAKCGKCFKHCITKXSPQDTYGO--DGMSSWTFEGS 563  
 DB 511 CTGSGSVLVLCQTKHFMADGTSCGSKWCTINKCNKTKDRKHFDTPFGSGMGMPPMD 570  
 QY 564 CSKSGGVSRSRSCNPSPAYGRCPLGPMFEYOVNCSSECP--GYEDFPAOCCAKRN 622  
 DB 571 CSPTCGGVOYTYMRCDNPPVKNKGKCYCEKRYKRSCHLECPDNNKGTFRBEDCEAHN 630  
 QY 623 SYVYHNAKHS-----WVP--YEPDDDAQKCELLCGSADTGDVYFNQVNHDTGRSY 673  
 DB 631 EF-----SKASFGSPRAVEWIPRYAGVSPKDRCKLICQAKGIGYFFVLQPKVVDGTPCS- 684

QY 674 RDPYVCARGECYVPCDCEKVEGSMKADKCGVCGGDNCHRTVGLTKASKOAGALKIV 733  
 DB 685 PDSTVYCGVCGCKKACDRIIDSKKKFKDCCGCGGNGSTCKKISSV--TSKRGYHDIT 742  
 QY 734 QIPAGRHQIETALEKSPHRS-----VKNQYTGSIPLMPKKEATSRFTAGLEWE--- 786  
 DB 743 TTPGTATNIEVORNRQGRSRRNNGSFLAIKADGYILN-----GDYTLSTLEQDLMY 794  
 QY 787 -----DAVEDAKSLTSGPLPEALIALPLPTEGEP--RSSIATYVYHEDLLPLIGS 838  
 DB 795 KGVLYRYSSSSALALRINSFSPLEPTLIQVL--TVGALPKIKITYFY----- 842  
 QY 839 NNVLLEMDTYEMALKSWAPCSKACGGCIQFTKYGCRRRRDHMYORHLCDHKRPKPIR 898  
 DB 843 -----KKK-----K 846  
 QY 899 RRCNQHPCQPYWYTEEMGACSRGKIGVOTRGIQCLPLPSNTHKVPKACACD--RP 957  
 DB 847 EEFNMLP--TFSAWVTEEWGECSSKSC--ELGWQRRLVEC--RDING--OPASECAKEVKP 898  
 QY 958 EARRPCLRPVCPAOWRLGAMSCSATCGEIQOQVQCRT--NANSLGHCSD--RP-DT 1012  
 DB 899 ASTRPCADHPCP--QWQJGEMSSCSKTGCKYKKSILACLSDHGVSLSHESCDPLKPKHP 957  
 QY 1013 VOVCSLPAC 1021  
 DB 958 IDECTMACC 966

RESULT 10  
 AT55\_HUMAN  
 ID AT55\_HUMAN STANDARD; PRT; 930 AA.  
 AC Q9UNAO; Q9UNP2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE ADAMTS-5 precursor (EC 3.4.24. .) (A disintegrin and metalloproteinase with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)  
 DE (ADMP-2) (ADAM-TS 11).  
 GN ADAMTS5 OR ADMP2 OR ADAMTS11.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=99367476; PubMed=10438522;  
 RA Abbaszade I., Liu R.-O., Yang F., Rosenfeld S.A., Ross O.H., Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M., Wynn R., Duke J.L., George H.J., Hillman M.C., Jr., Murphy K., Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H., Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F., Arner E.C., Burn T.C.;  
 RA "Cloning and characterization of ADAMTS1, an aggrecanase from the ADAMTS family";  
 RT J. Biol. Chem. 274:23443-23450(1999).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20289799; PubMed=10830953;  
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E., Ohtsuki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W., Rosenthal A., Sasaki T., Shibuya K., Kawasaki K., Asakawa S., Shimizu S., Shimizu N., Nordliek G., Hornischer K., Brandt P., Schafte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H., Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehnach H., Reinhardt R., Yaspo M.-L.;

FT	DOMAIN	257	261	POLY ARG.	..	(POTENTIAL).
FT	CARBOHYD	498	498	N-LINKED (GLCNAc.	..)	(POTENTIAL).
FT	CARBOHYD	728	728	N-LINKED (GLCNAc.	..)	(POTENTIAL).
FT	CARBOHYD	802	802	N-LINKED (GLCNAc.	..)	(POTENTIAL).
FT	CARBOHYD	807	807	N-LINKED (GLCNAc.	..)	(POTENTIAL).
FT	CONFLICT	138	138	A -> G (IN REF. 2).		
FT	CONFLICT	614	614	R -> H (IN REF. 3).		
FT	CONFLICT	692	692	P -> L (IN REF. 2).		
50	SEQUENCE	930 AA:	101715 MM:	B6428150F28193B	CRC64:	

```
Query Match      15.4%; Score 1068; DB 1; Length 930;
Best Local Similarity 30.0%; Pred. No. 1.4e-48;
Matches 309; Conservative 119; Mismatches 362; Indels 240; Gaps 43
```

Qy	59	GPAASAGSNVDTPEPL-----PRHSHLRV---ARSPH-----	92
Db	21	GPAATPADCK-AGOPPTAAAAAPRRKQGEVEVEREPPCHPHPLAQRRRSGLYONIDQ	79
Qy	93	---GGTLMPGVGVGHSHLSYENVYAGKEKLRLRPNRLL-----VVP---GSSYEWODEFRE	142
Db	80	LYSGS---GKVG-----YLYVAGRGFFLLDLERDSVGLAGVVPARGGTSAPWRH----	126
Qy	143	LEPROLROECYVYTGVTGMPGAAVAISNCDLGLLRTDSTDEFFLEPLRGQOEKASCR	202
Db	127	-----RSHCHYRCGVADASPRSLAVFDLCGLDGFVAFKHARTLEPLLRGMABEEKGR	180
Qy	203	TH-----VYKREAVQDEMAEPDD-----DLHNEAGLGLDPLNLGLVODOLG	245
Db	181	YAGDSARILHYVYREGFSFALPRPASCETPASTPRAHHAHAHSH-DSGRAALLSOLL	239
Qy	246	DRE-----RKRRHAKPGASYSEIVLLVVDSDVYVRHKGHEVYONVLTMINIV	291
Db	240	DQSALSPAGSGSPQTMWRRRRRRSISRARQVDELLVADASARILXR-GLQHYLTLASTA	298
Qy	292	DEIYHESLGVHINAIYVRLIMVGYROSLILIERGNPSRSLBOVCYRMAHSQORODPSHA	351
Db	299	NKLYSHASILEHNIIRLAVYKVVYVLDGDK-SELYSKNAATLLKFKCKMOHNOGLDDBE	357
Qy	352	HHDVVELTRQDF-----GRSGVAPYMGCHPLRSLALHDEGFSAPYIAHETGHVIG	405
Db	358	HYDAAILFTFRDELCSHSCDTLGMADVGTICSPRSCAVIEDDBLAATVAHIEIHLIG	417
Qy	406	MEDHOGMGCADETSLGS-----VMAPIVQAARFHRFMSRCSKLEISRYLPS---YDCLLD	458
Db	418	LSHD-DSKFC--EEFFGSTEDEKRLMSSILSISIDASKFMKSQTSATITFELDDCHGNCLLD	474
Qy	459	DFPDPAPOPPLPEINTSMDOCHFDFGSGIOTCLAFRFEPRKQJWMS-HPDNFYFCK	517
Db	475	LBRKQIL-GBEPLPQOTYDATQOCCULTGPEPYSVCPG--MDVCARIMCAVVRQOGWCL	530
Qy	518	TKKGPRLDEBPACPKMCFKGHCIMKSPQRY--GODGMSWPKFEGSCSRSGGGVRSR	575
Db	531	TKKLPAVESTPCGKGKRICLQCKYDCTKKKYYISHSNMGMSWMSWGQSRSGGGVQFA	590
Qy	576	SRSNNPSPAYGGRPLCGPMEFYOVNSSECPGTYEDFRAOQCAKRNSTYVHONAKH----	632
Db	591	YHRCNPNAPRNNRGVCTGKRAIYRSCSILMPCBPMPKSFRRHOCFAKNGY--QSDAGVKT	648
Qy	633	--SWP-YEPDDDAQKCELTQCSADTGGVYFMNOVYVHGFRCFSRDY--SVARGECVP	687
Db	649	PVEMVKEYVGLPADVCKLTCTRAKGTGYVVSPEYVDGIECR--PYNSVCYKRGCVR	705
Qy	688	VGGDKFVGAMKADKCGVCGGDNCSRCPIVKTSLTGAKSKOAGALKVLIOPAGARIQLE-	745
Db	706	TGCDGILIGSKLOYDKCGVCGDNSSCTKIYGFNNKKS--GYDVVRLPEAATHIKVROF	763
Qy	746	-----ALEKSPHRSVYVKNQVTSFTLNFKEGKEATSRF-----TAMLEMEDA	788
Db	764	KAKDQRTTAYALAKK-----KN-----GEYLLINKYMIWSTETIINDINGVNNYSGWSH	813
Qy	789	VED-----AKESLKTSGPLPEAIINALPPEEGSPRSLSIYKVYIHEDLLPLGSN	839

```

Db 814 RDDFLHGMGYSATKEIL-----IVQILANDPTK---PLDVRYSEFVPKKSTPKVNS- 861
QY 840 NVLLEEMDTYEWALKSMACSKACGGGIOFTKYGCCRDRDHMMVQRLCDHKRKPPIRR 899
Db 862 -----VTSHGSKNKVGSN----- 873
QY 900 RCNOHPCSQPVVTEBEGACSRSCGLGVOTRGIOCLPLSNGTHKYMPAKAC-AGDRPE 958
Db 874 -----TSQPKWVTGPMILACSRIC-DTGMHTRTVDC-----QDGNKRL--AKGCPLSQRP 920
QY 959 ARRPCLRVPC 968
Db 921 AFKOCCLKKC 930

```

Job completed: August 28, 2002, 20:14:54  
 Elapsed time: 261 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: August 28, 2002, 19:52:08 ; Search time 66.13 Seconds  
(without alignments)  
2102.894 Million cell updates/sec

Title: US-09-938-330-20  
Perfect score: 6954

Sequence: 1 MAPLRALLSYLPLHCACT.....PTGTFLLCVLPRDSQLRGHT 1252

ing table: BLOSUM62

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%

Existing first 45 summaries

Database :

1. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1980.DAT \**  
2. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1981.DAT \**  
3. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1982.DAT \**  
4. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1983.DAT \**  
5. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1984.DAT \**  
6. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1985.DAT \**  
7. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1986.DAT \**  
8. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1987.DAT \**  
9. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1988.DAT \**  
10. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1989.DAT \**  
11. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1990.DAT \**  
12. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1991.DAT \**  
13. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1992.DAT \**  
14. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1993.DAT \**  
15. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1994.DAT \**  
16. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1995.DAT \**  
17. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1996.DAT \**  
18. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1997.DAT \**  
19. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1998.DAT \**  
20. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA1999.DAT \**  
21. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA2000.DAT \**  
22. *SlDS5/gcgdata/geneseq/geneseq-emb1/AA2001.DAT \**

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5160.5	74.2	1048	22	AA856955	Larynx carcinoma
2	3624	52.1	1205	22	AA873550	Human ADAM-type me
3	3614	52.0	1201	21	AA821254	Human metallopeptid
4	3327	47.8	1203	22	AA850004	Bovine metallopeptid
5	3299.5	47.4	1211	19	AA847028	Human N-proteinase
6	3279.5	47.2	1338	22	ABG22373	Novel human diageno
7	3270.5	47.0	1206	19	AAW47030	Bovine N-proteinase
8	2962.5	42.6	969	21	AAV53900	Amino acid sequenc
9	1647	23.7	566	19	AAW47029	Human N-proteinase
10	1438	20.7	1586	22	AA874944	Human ADAM-type me
11	1436	20.6	1686	22	AAE00913	Human 27875 ADAM-T

12	1435	20.6	1668	22	AAE00934	Human	27875 ADAM-7
13	1436	20.5	1690	22	AAE86949	Human	metalloprotease
14	1415	20.3	1784	21	AAB41379	Human	ORFX ORF1143
15	1328	19.1	1103	22	AAB74945	Human	ADAM type metal
16	1240.5	17.8	1193	22	AAB72301	Human	ADAMTS-9 alt
17	1221	17.6	1072	22	AAB72300	Human	ADAMTS-10 al
18	1220.5	17.6	1081	22	AAB72288	Human	ADAMTS-10 am
19	1216	17.5	1073	21	AAB21264	Human	metalloprotease
20	1164.5	16.7	959	22	AAE86947	Human	metalloprotease
21	1159.5	16.7	908	22	AAE03572	Human	metalloprotease
22	1147.5	16.5	1882	21	AAB72286	Human	protease-rel
23	1121.5	16.1	958	21	AAB71255	Human	ADAMTS-9 am
24	1118	16.1	950	22	AAE62299	Human	metalloprotease
25	1117.5	16.1	874	22	AAB72287	Human	metalloprotease
26	1110.5	16.0	947	22	AAB68930	Human	ADAMTS-9 am
27	1102.5	15.9	968	22	AAE50011	Human	metalloprotease
28	1098.5	15.8	950	20	AAB49501	Human	metalloprotease
29	1098.5	15.8	950	22	AAB73549	Human	prol; SER ID 12
30	1098.5	15.8	950	22	AAE50002	Human	ADAM-1 protei
31	1098.5	15.8	967	19	AAE08285	Human	ADAM-type me
32	1098.5	15.8	967	20	AAE04182	Human	integrin lig
33	1088.5	15.7	967	20	AAE04189	Human	Tango-71 pro
34	1079.5	15.5	950	21	AAE33899	Human	secreted pro
35	1075	15.5	2150	21	AAE33898	Human	acid sequenc
36	1072.5	15.4	930	22	AAE05617	Human	secreted pro
37	1068	15.4	930	20	AAE72283	Human	aggreant de
38	1053.5	15.1	997	22	AAE10350	Human	ADAMTS-7 am
39	1052.5	15.1	891	22	AAE10350	Human	ADAMTS-7 am
40	1052.5	15.1	929	21	AAB41236	Human	ADAMTS-11.4
41	1040	15.0	859	22	AAB72282	Human	ORFX ORP990
42	1038.5	14.9	1054	22	AAB60410	Human	ADAMTS-6 am
43	1037.5	14.9	757	22	AAE03353	Human	ADAMTS-10 am
44	1032	14.8	896	22	AAB71255	Human	protease-rel
45	1025	14.7	930	22	AAE71280	Human	ADAMTS-5 am

## ALIGNMENTS

PT	PN	RESULT
XX	XX	1
XX	XX	AAB85695
ID	XX	AAB85695 standard; Protein; 1048 AA.
XX	AC	AAB85695;
XX	DT	29-OCT-2001 (first entry)
XX	DE	Larynx carcinoma associated protein-1 (larcap-1) polypeptide.
XX	KW	Larynx carcinoma associated protein-1; larcap-1; carcinoma; cytostatic; antineoplastic; antirheumatic; osteopathic; immunosuppressive; human; immunostimulant; cerebroprotective; vasotropic; antiinflammatory; dermatological; cardiac; vaccine.
XX	OS	Homo sapiens.
XX	PN	WO200159133-A1
XX	PD	16-AUG-2001.
XX	PF	12-FEB-2001; 2001WO-EP01525.
XX	PR	14-FEB-2000; 2000EP-0102955.
XX	PA	(MERE ) MERCK PATENT GMBH.
XX	PI	Duecker K, Hentsch B, Hoheisel J, Frohme M;
XX	DR	WPI: 1999-483569/52.
XX	DR	N-PDB: AAH47048.
XX	XX	Larynx carcinoma associated protein-1 polypeptide for the treatment of carcinomas, metastasis, arthritis, osteoporosis, immune disorders,

PT stroke, ischemia, autoimmune disease, angiogenesis, skin disorders and  
PT organ malfunctions -

XX Claim 1: Page 33-39; 47pp; English.

XX This represents a human larynx carcinoma associated protein-1 (larcap-1)  
CC polypeptide. The larcap-1 polypeptide can be expressed by standard  
CC recombinant methodology. The larcap-1 polypeptide and polynucleotide are  
CC useful in diagnostic assays and for the treatment of carcinomas,  
CC metastasis, arthritis, osteoporosis, immune disorders, stroke, ischemia,  
CC autoimmune disease, angiogenesis, skin disorders and organ malfunctions  
CC especially heart hypertrophy.

XX Sequence 1048 AA:

Query Match 74.2%; Score 5160.5; DB 22; Length 1048;

Best Local Similarity 95.4%; Pred. No. 0;

Matches 925; Conservative 0; Mismatches 6; Indels 39; Gaps 2;

291 VDEIYHDSIGVHNTALVRLIMVGYRSLIERGNPSRLERQVCRMAHSDQRODPSHA 350

80 vdeiyhdesigvhnalvrlimvgyrslsiergnpsrslieqvrwahsqgrqdsna 139

351 EHHBHVFLTRDGFSGYAPVTGMCHPLRSALNHEDGSSAFVIAHETGHVLMEDHG 410

140 ehbhvfltrdgtfsgyapvtgmchplrsalnhdgssafviahetghvlgmedhg 199

411 OGNCACBETISGYMAPLNOAAPHFHMSSKLELSRYLPVSDCLDDDFDPAWPOPE 470

200 qngcacetisgymaplvgaalnrafhfmskslelsrylpvscldddfdpawpope 259

471 LPGINSWDEOCRFDFSGYQTLAFRTFPCQQLMCSHDPNPFCKTKKPPLDGTECA 530

260 lpginswdeocrdfdfsgyqtlafrtfpcqqlmcsdpnpfcktkkppldgteca 319

531 PGKMCFFKHCILWKSPEQTYGDDGMSWTKFGSCRSRSGGVRSRSCNNPSPAVGGR 590

320 pgkmcffkchcilmkspetlygddgmswtkfgscrsrsggvrstrscnnpspavgr 379

591 CLGPFMEYQVCSSECEGTEDFRAQOOCARNSYVHQNKHSMWPEPDDDAOKCELIC 650

380 clgpfmeyqvcsseceptyedfraqoocakrnsyvhqnkhsmwpepdddaokcelic 439

651 QASDTGVFVNNVYHDCSTGCRSDPYSVCAKRGCVPGCDKEGSMKADKCGVGGDN 710

440 qasdtgdfvnnvyhdcstgcrsydpsvcargcvpvgcdkevsmkaddkcgvggdn 499

711 SHCRVKTGLKASKOAGALKLIVQIPAGARHIOLEALEKSPRSVKNNOVTSFILNPKG 770

500 shcrvktglkaskoagalklvqipagarihiolealeksprrsvknnovtgspfilnpg 559

771 KEATSRFTFAMGLEWEDAVEDAKESLKTSGPLPAIALALPREGGRSSLAKEYIHE 830

560 keatsrftfamlglewavedakeslktsgplpaialalpreeggrsslakeyihe 619

831 DLPLIGSNVYLEMDYEWALKSMAPCSKACGGGIDFTYGGRRRRDHMMVORHLCDH 890

620 dlpligsnvylemdyewalkswapcskaaggidftyygrrrrdhmmvgrhldch 679

891 KKRPRPIRRRCNOHPCOSPVWTEEMGACSRSCGLVGTGRIQCLPLSLNGTHKWPAP 950

680 kkrprpirrrcnohpcospvwteemgacsrscglvgtgriqclplslngthkwpap 739

951 ACAGRPFAKRRCLAVPCPAOMRLGANSO----- 979

740 acagdrpearipclrvpdpagvrlgawsgkylstscmpdlvlmrepsingteliallv 799

980 -----CSATCGEGIOOROVVCRRTNANSLGHCEDRPTVOYCSLPACGHNQNSTVRADV 1034

800 qptvlscatcgsglqgrvcrvtananslghecdgdrptdvtvcslpacggnbnqnstvradv 859

1035 WELGTPEGQWVPOSGPLHPIKINISSMCAABPCGTGRSVFCOMEVLDRCISIPGHRLCV 1094

DB ||||| 860 welgtpeqwwpseghpblnkiss---tepctgdsivfcgmeyldtyscipyhnlccv 916

QY 1095 SCIKKAGPNPDPGPPTSLPPEFSTPGSPLPGDPPADAAEPGKPTGSDHOGRAIOL 1154

DB 917 scikkaagpnppdpptslpfpstpgsplpgdpdpadaaepgkptgsedhgratql 976

QY 1155 PGALDTSSPGTOHPFAPETPIPCASWSISPTTGGCLFMGWTOTPTVPEDKGGPGLRH 1214

DB 977 pgaldtsspgtqhpafetpdpigawswispttggclfmgwtotptvpedkqgpgdelrh 1036

QY 1215 PGTSPLADLP 1224

DB 1037 pgtslpaasp 1046

RESULT 2

AAB73550

ID AAB73550 standard; Protein: 1205 AA.

XX AAB73550;

XX 07-AUG-2001 (first entry)

DE Human ADAM-type metalloprotease MDT5, SEQ ID NO:10.

XX Human, MDT5; ADAM-type metalloprotease; drug screening;

XX A Distintegrin And Metalloprotease; cancer; arthritis.

OS Homo sapiens.

XX JP2001017183-A.

XX 23-JAN-2001.

XX 09-JUL-1999; 99JP-0196584.

XX 09-JUL-1999; 99JP-0196584.

XX (YAMA ) YAMANOCHI PHARM CO LTD.

XX WPI, 2001-275950/29.

XX N-PSDB; AAB70225.

XX A new metal protease and its preparation for use as an anti-cancer and

XX anti-arthritis therapeutic -

XX Claim 1: Page 16-19; 22pp; Japanese.

XX The invention relates to the novel human ADAM (A Disintegrin And

XX Metalloprotease)-type metalloproteases MDT5 (AAB73549) and MDT5

XX (AAB73550). The metalloproteases can be used for the treatment of

XX cancers and arthritis. The invention also relates to the genes encoding

XX MDT5 and MDT5, vectors and host cells containing the MDT5 or MDT5

XX genes, the recombinant production of MDT5 and MDT5, and antibody

XX specific for MDT5 or MDT5, and methods of screening for compounds

XX which modulate the activity of MDT5 and/or MDT5. The present

XX sequence represents human MDT5.

XX Sequence 1205 AA;

Query Match 52.1%; Score 3624; DB 22; Length 1205;

Best Local Similarity 55.6%; Pred. No. 2.8e-249;

Matches 666; Conservative 145; Mismatches 233; Indels 152; Gaps 17;

QY 37 DYGTVPCSTDFRGRFLSHVSGPAAASAGSMVVDPTLPRHSRLR-----VARSPLH 91

DB 41 eyelivpvslnlgrylshls-----ashktrsdavsnp-- 77

QY 92 PGGLMFGVGRHSILYFNVTVEGKEHLRLRPNKRLVFGSSVEMOE----- 138

DB 78 -----eqllfnltafgkdfnlrlkplqtlvapgavvewhetaivpgnltcpin 125

```

OY 139 -----DERELFROPLROECVYTGVTGMPGAVALNSMCDLGLIRTDSTDFEIPLE 191
Db 126 nmgpsatyrirktleplqncayvgdivdipgtsvalnsncdglagmlksdneyle 185
OY 192 RGOOEKASGRTHVYVYRREAVOEAEPDGLH--NEAFGLDLPNLLGLVGDQIDGTE 248
Db 186 rkgmeeeegrthvykrsavegapldmskdfhyresdlegldlgytynghqnlctm 245
OY 249 RKRRAKPGSIEVLLVAVDSSVRFHKGKHQNVLLMLNVDIYHDESLGVHINITAL 308
Db 246 rrrrhagendynlelvlygddsvvrfhkgheqnylltlimlveinydeslgyvlnv 305
OY 309 VRLVWGYROSLLIERGNPSRSLERQVMAHSQORODPSHAENHDHVVFLRODFGS- 367
Db 306 vrmnlgyaksisltergnpsrslenvrwasqgrslmnehdhalfiltrqdfpbg 365
OY 368 --GYAPVYGMCPRLRSCALNHEDGSSAFVIAHETGVLGMEHDGNGCADETSIGSV 425
Db 366 mgyavpvtgmchprvscldnhdgfsafviahetghvlgmehdgqgnrcgdetamgs 425
OY 426 APVVOAFHFRHMSRCSKLELSRYLPSYDCLLDPEFDPAWQPELRLGINSMBQCFD 485
Db 426 aplvvaafhryhmsrscsgelkrylshydclddpfdhdpkldpqlpslmsdegcrfd 485
OY 486 FGSGYOTLAFRTEPCQOLWCSPNDNPFCKTKKGPRLDTECAPGKWCFCGHCIMKSP 545
Db 486 fgvykmtcalfitfpdqqlwscshndpnyfcktkkqprldtecaagkwcgkymcmwka 545
OY 546 EOTYODGNGSSWTKFGSCSRSCGCVARSRSRSCNNESPAYGPRCLCPMEFYOVCSSE 605
Db 546 ng-qkqdgngswtkfgscsrscgcvarsrsrscnnespaygprclcpmefyovcsse 604
OY 606 CPGETEDFRACQACAKRNSYVYHONAKSMWPEBDDAOKELLCOSADNDGVYFMNQV 665
Db 605 cqhtfedfraqqcgqrshfeyqntkhwlypndpdkrchlycgskedgvaaymqvlv 664
OY 666 HDGRCSRDRPYSVACARECVVSCDKEVSGMADKCGVCGGNSHCRTRKGLGKASK 725
Db 665 hdgthcsydkpyslcvrgecvkvgcdkeigsnkvedkcyvgngshncrtvkgfctpr 724
OY 726 OAGALKLVQIPAGARHIOIELESPIHRSVYKNQVTSFLLNPKGKATSTFTFAMGLEW 785
Db 725 klyglkmdlppgarhviqgedesphlalknqatbhyllngkgeaakstfildigvew 784
OY 786 EDAVEDAKESLKTSGPPEAIAIALPPTREGPRSSLAAYVYHEDLLPLIGSNVLEE 845
Db 785 dnyteddieslhtdgrhdpviviil-pgendtrsslttyyilhedsvplinsnnvlee 843
OY 846 MDYEMALKSNAPCSKACGGGIGTGTGCRRRDHVWORHLCDHKRRPRIRRCNQHP 905
Db 844 ldtlewalikswsgskpcqgfytkygcrrksdkmwhsfceankkprfrimnige 903
OY 906 CSOPVWTEEMGACSRSCGLGVOTRIGIOCLPLSNGTHVMAPKACAGRRPARPCLR 965
Db 904 cltpvlwaeeewenctkcgssgyqlitvrcldplldgtnsvnskcmgdrpessrpn 963
OY 966 VPCPAQRILCAWSOCATCGEGIQOROVYCRTNANSLGHCEDRPTVYVCSIPACGNGH 1025
Db 964 vpcpaqkltgvrwesecvtecgtegrvqlcrag----dhcdgkcpesvtaeqclpnd-- 1017
OY 1026 QNSTVRADVMELETPBEOQWPOSGPLHPINKISSMCAERCTDRSVFCOMEVLDRYCST 1085
Db 1018 -----epclgdkslfcqmevlaryst 1039
OY 1086 PGYHRLCCVSCIRKASG-----PNMGPPPG-----PTSLPPF--STPG- 1121
Db 1040 pyynklccescskrstllpppyllaeaetdvaisnpsdprslvymptslvphsctepk 1099
OY 1122 -----SPLPQPDADAAEPKCPMGSGEDHOGHATQILPGALDT-----SSPQTQ 1166
Db 1100 kmslssissvgnpaya-alfpnskpddganlrg--rsaqgagsktvlrvltvpspsptk 1154

```

```

RESULT 3
AAB21254
ID AAB21254 standard; Protein; 1201 AA.
XX
AC AAB21254;
XX
DT 23-FEB-2001 (first entry)
XX
DE Human metalloproteinase KIAA0366.
XX
KW Human; KIAA0366; ADAMTS; metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; neurotrophic; neuroprotective; antiparkinsonian;
KW cerebropoietic; cytosolic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
XX
OS Homo sapiens.
XX
PN W0200053774-A2.
XX
PD 14-SEP-2000.
XX
PE 08-MAR-2000; 2000MO-US06237.
XX
PR 08-MAR-1999; 9905-0264585.
XX
PA (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
PI Kelnner GS, Clark M, Maki RA;
XX
DR WPI; 2000-594326/56.
XX
DR N-PSDB; AAA95824.
XX
PT Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX
PS Claim 13; Fig 8; 129pp; English.
XX
CC The present sequence is human metalloproteinase KIAA0366. It
CC shows homology to the ADAMTS family of proteins, which contain
CC thrombospondin, integrin and metalloproteinase domains. ADAMTS
CC polypeptides are useful for the manufacture of medicaments for treating
CC conditions associated with neuroinflammation and/or neurodegeneration,
CC such as Alzheimer's disease, Parkinson's disease and stroke. They are
CC also useful for treating conditions associated with cell proliferation,
CC cell migration, inflammation and/or angiogenesis, such as cancer,
CC arthritis and autoimmune diseases. They can be used to treat patients
CC afflicted with an invasive tumour, a brain tumour or brain injury.
XX
SQ Sequence 1201 AA;
XX
Query Match 52.0%; Score 3614; DB 21; Length 1201;
Best Local Similarity 55.5%; Pred. No. 1.5e-248;
Matches 665; Conservative 145; Mismatches 236; Indels 152; Gaps 17;
OY 37 DYGVTVPCTDPRGRFLSHVSGPAAASGSMWVDTPTLRHSHSLR-----VARSPLH 91
Db 37 eyelvtpvstnlegrylshls-----ashkrrsardvsnp-- 73
OY 92 PEGTLPMPGVGRHSLYFVNTVYGRKELHLRLPRNRLVYVPGSSVEMQF----- 138
Db 74 -----egltfntafgkafhlrlkpnqvlavagavvwehelslypnaitpin 121
OY 139 -----DERELFROPLROECVYTGVTGMPGAVALNSMCDLGLIRTDSTDFEIPLE 191
Db 122 nhgpsatyrirktleplqncayvgdivdipgtsvalnsncdglagmlksdneyle 181
OY 192 RGOOEKASGRTHVYVYRREAVOEAEPDGLH--NEAFGLDLPNLLGLVGDQIDGTE 248

```



Matches 641: Conservative 160; Mismatches 351; Indels 76; Gaps 20:

```

OY 7 LLSYLLPHICALTAAGSRPEELHLSKLDYGVTVCCSTDFGRRLSHVSCPAASAAG 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17 |||||lp|padar|laaaadppgpgqgaaellavpvrtdaggrlvshvsa-atapag 75
OY 67 SMVVDTPPLPRHSHLRVAKSPPLHPCGTLMPGRVGRSHSLYFVWVYEGKELRLRPNR 126
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 76 ---vrrtrapaqipjls-gsseedpgs-----rlfyvvtvfgtdlhlrlfllpar 121
OY 127 LVVGGSSVEMWEDFRELFRQPLROECYVTCGVGM-PGAVAISNCDGLAGLRIDSTDP 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 lvapgatvewgseagatrvp|lgtcl|yvgdvag|laessvalnsodg|agllimeeef 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 FIEPLERGOOEKKA-SGRTHVYRREAVQOEMAPDG---DLHNEAFGLDLPNLLGLVG 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 ftepleglaakeeggrvnyvnyhrrtrtr--ppplqgaldtgisadsislsrlygie 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 242 DQLODTERK-RRAAKPGSSYSIEVLLVVDSSVFRHKEHQNVTLMNIVDEIYHDESL 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 ervnsrrtrrhaaddynleavl|gvdsdvqfng|tenvqy|l|lmlivei|yhesl 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 301 GVHNINLALVRLWGYRQSLIERGNPSRLDQVCRMAHSQQRDPDPAEHHDHVVPLT 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 gahlnvtrvl|llsygsm|lelignpsq|lenvcrrway|lqk|p|td|hey|h|ha|f|l|t 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 361 RQDFGPS---GYAPVTGMCPLRSCALNHEDGFSAFVAHENGVLGHEHGGCGNGCAD 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 rgtlqpsgmgyapvpmgmchprvsc|l|n|hed|f|s|a|f|v|a|h|e|g|n|v|l|g|h|e|g|q|n|r|c|g|d 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 418 ETSIGSMALVQAARHRRHMSKSLERLYLPSYDCLLDPPDPAPQPELPGINYS 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 evlrgsmabl|vqa|a|f|h|r|w|s|r|c|s|q|e|l|s|y|l|h|s|y|d|c|l|r|d|p|f|h|d|w|a|l|p|q|l|n|y|s 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 478 MDEQCFEDFSGGYOTCLARFTEPCQKOLMCSHPDNPFECKTKKSPRLDGTBAPAKMCFK 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 mneqcfidf|g|y|m|m|c|t|a|f|t|f|p|c|q|l|w|s|h|p|d|n|y|t|c|k|k|p|p|l|q|t|m|c|a|p|h|k|h|c 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 538 GRGILKSPEDTGYQDGGSSWTKPFGSCSRSGGGVRSRSRSCNNSPAYGGRPLCGPFE 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 540 ghciw|l|p|d-|l|k|r|d|g|n|w|a|s|p|f|g|s|r|c|t|q|l|y|k|f|r|t|q|d|h|p|n|g|r|t|c|s|g|l|a|y|d 598
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 598 YQVGNSEECGFYEDFRQAQCAKRNQSYVHONAKHSWVPERDDAQCCELICQADTGD 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 fqlcnsqdc|p|d|a|t|t|e|e|q|c|r|q|w|d|l|e|h|g|d|a|h|w|l|h|e|h|r|d|e|k|e|h|c|h|y|c|s|k|e|t|e|g|e 658
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 658 VPFMANOVHDGTRGYSRDPYSVARGECVPYGCDEKESVMAKADKCGVCGGDNISGCRPVK 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 659 vvmkrmvhdgtrcsykdafslcvrgdc|k|y|g|c|d|v|s|s|k|g|e|d|c|y|g|c|g|a|h|c|k|v|k 718
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 718 CTLGKASKAGALKLVQIPAGARHIQIEALEKSPHRSVYKQVTSFLNPKG-EATSR 776
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 719 gtfstzspk|l|y|k|f|e|l|p|a|g|a|h|l|l|q|e|d|t|s|h|l|a|k|h|e|t|g|k|f|l|n|e|e|n|d|v|p|n|k 778
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 777 TTFAMGLEMEDAVDAKESLTKTSGPLPEAIIALPTEGPRSSLAKEYIHEDLLPLI 836
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 779 ttfamglemedavda|k|e|s|l|t|k|t|s|g|p|l|e|a|i|i|a|l|p|t|e|g|p|r|s|s|l|a|k|e|y|i|h|e|d|l|l|p|i 835
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 837 GSNVILVEMDTYEWALKSWAPCSACGGG|Q|FTKYGCRRRRDHMYQRIHLCDKKRPK 896
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 836 ddnn|l|e|d|s|v|g|e|w|a|l|k|k|w|s|p|c|s|k|p|c|g|s|g|f|k|y|g|r|t|l|h|k|h|w|h|r|f|c|d|s|v|s|p|k|a 895
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 897 IRRCONHRCSPVWVTEEMKAGSRSCCKLGVOTRG|Q|CLPLSNGTHKWPAPAKACADR 956
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 896 irrconhrcspv|w|v|t|e|e|m|k|a|g|s|r|s|c|c|k|l|g|v|o|t|r|g|q|c|l|p|l|s|n|g|h|k|w|p|a|p|a|k|a|c|a|d|r 955
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 957 PPARPPCLRVPCPAQWRLGANSOCASATCGEIQOQVVCFRNANSLGCHGSDRDPDYQVC 1016
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 956 pparppclrv|p|c|p|a|q|w|r|l|g|a|n|s|o|c|a|s|a|t|c|g|e|i|q|o|q|v|v|c|f|r|n|a|n|s|l|g|h|c|h|g|s|d|r|p|d|y|q|v|c 1015
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1017 SLPACGNGHONSTVADVWELGTPGCGVVPQSGPLPIPKISSMCAAEPTGDRSVFCOM 1076
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1016 rlgpcrntsdpskksy-----vgw|s|r|p|d|n|s|p|v|g|e|t|s|s-----k|g|r|c|g|d|k|s|v|f|c|m 1065
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

OY 1077 EYLDKYSIPCYHNLCCVSC-----IKKASGPNCPDPTGPTSLPFPSTPGSPPL---- 1124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1066 evlrsy|s|p|g|y|n|k|l|c|k|s|c|m|h|n|l|d|v|d|d|t|e|p|p|g|h|n|d|e|e|l|m|p|l|s|p|l|y|m|e|w|q 1125
OY 1125 PGPDDPADAAEPPEKPPGSEDHOGRAVTLPGALDTSSPCTQHFAEPETPGASWSISIP 1184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1126 ppp|p|l|e|v|p|n|t|s|t|n|a|t|e|h|p|e|t|n|a|v|d|y|k|i|-----p|g|l|e|d|e|v|g|p|l|l|p|----- 1173
OY 1185 TTPGCLPMGWTQPTPTVPEEDKGGPGEGL 1212
Db 1174 -----r|p|s|p|y|e|k|t|r|n|g|r|i|q|e|l 1190

```

RESULT 5  
AAW47028  
ID AAW47028 standard; protein; 1211 AA.  
XX  
AC AAW47028;  
DT 06-JUL-1998 (first entry)  
DE Human N-proteinase (130 kDa long form).  
XX  
KW N-proteinase; human; collagen; antibody; rheumatoid arthritis;  
KW fibrosis; Ehlers-Danlos disease; diagnosis; therapy.  
XX  
OS Homo sapiens.  
XX  
PN WC9800555-A1.  
XX  
PD 08-JAN-1998.  
XX  
PE 03-JUL-1997; 97WO-0512427.  
XX  
PF 02-JUL-1997; 97US-0886333.  
PR 03-JUL-1996; 96US-0021203.  
XX  
PA (COLI/) COLIGE A.  
PA (LARI/) LAPIERE C.  
PA (PROC/) PROCKOP D J.  
XX  
PI Colige A, Lapiere C, Prockop DJ;  
PI WPI: 1998-086980/08.  
DR N-PSDB: AAY06592.  
XX  
PT Polynucleotide sequence encoding human N-proteinase - used to  
PT produce mature collagen in vitro and antibodies to treat fibrosis  
PT and rheumatoid arthritis  
XX  
PS Disclosure: Fig 1D; 49pp; English.  
XX  
CC This polypeptide comprises the 130 kDa long form of human  
CC N-proteinase, the enzyme responsible for cleaving N-propeptide from  
CC procollagen to produce mature collagen. The amino acid sequence  
CC was deduced from a cDNA sequence (see AAY06592) derived from human  
CC skin fibroblast cDNA clones. The 70 kDa short form (see AAW47029) of  
CC human N-proteinase was also identified. Human N-proteinases  
CC can be produced using a claimed method in which a host cell is  
CC transformed or transfected with an N-proteinase polynucleotide  
CC sequence, cultured in an appropriate culture medium, and the  
CC N-proteinase is isolated from the medium. The N-proteinase can be  
CC used for the production of mature collagen in vitro and for the  
CC production of antibodies which may be used for diagnosis and  
CC therapy of diseases including fibrosis and rheumatoid arthritis.  
CC The N-proteinase may also be administered to treat a disease  
CC resulting from insufficient production of N-proteinase, such as  
CC Ehlers-Danlos disease.  
XX  
SQ Sequence 1211 AA;

Query Match 47.4%, Score 3299.5, DB 19, Length 1211;  
 Best Local Similarity 52.6%, Pred. No. 3.8e-226;  
 Matches 635; Conservative 157; Mismatches 352; Indels 63; Gaps 20;

1 MAPRLALSTYL-----LHCACTAGAGSRTPEHLHSGKLSDVGVTPCSTDRG 50  
 12 lcpallllllllppllppppparlaaaadppgprl---ghaaellavprtdagq 68  
 51 RELSHVSGPAASAGSMVVDPTPLPRHSHLRVARSPLMHGTLMPGRVGRSHLYRNV 110  
 69 lrvshvsa-astragvrrraap-----vtrpsfpgn--eepgsh-llfynv 113  
 111 TVFGKELRLRLRNRLVVPSSVEWQEDFRLFRQECVYRGVGMGA-AVALIS 169  
 114 tvfgrelhlrlrparlvapga tmeqgkyltrvepllgsclyvgdvaaglaeasvals 173  
 170 NCDGLAGLRITDSTDFEIRERGOEKA-SGRTHVYRRARAVQOEAMEPDG-DLHNEA 227  
 174 ncdglagllrimeeeifteplekylaageagrvhvyrrtrptprlgppqaldtgaal 233  
 228 FGIGDLPLNLGIVDQLGDTERR-RRHARPGSYLIEVLLVDDSVYREHGEKHYONYLT 286  
 234 dsldslstralgveehansrrrrhaaddynlevllyvddsvvqfngkehvkylt 293  
 287 LAMIVDEIYHDESLGVHINIALVRLIMGYRQSLIIEGNPSRSLIEVCRAHNSQROD 346  
 294 lmiivdeiyhdeslghahlvrlilllsgykmsslliegnpsqlenvcwaaylqgkpd 353  
 347 PSHAEHHDVVFLLRQDFGSPS--GYAPYGMCHPLRSCALNHEDGFSSAFVIAHETGV 403  
 354 tghdeyhdhailftrtdfgsgmggyaprtgmchprscrlnhedgfssafviahelgv 413  
 404 LGMEHDGQNGCADFSLGVAAPLYQAARHNRHNRSCSKLESLRYLRYDCLLDPEPD 463  
 414 lgmehdgqngrcgdevr lgsimavlvqaathrwarscqselsrlyhndclldpbfh 473  
 464 AMQOPRELPGINYSMDQCFDFGSGYQCLAFRTFEPCKOLMCSHPDMPYFCKTKKGP 523  
 474 dmpalpqlpjlhysmeqcfdfglgymmtatftrtdcqlwcpndpnyfcktkkgrp 533  
 524 LDGTECAPGWKCFKGCIMKSPQOTYGOOGWSSWTKFSGCSGCGVRSRSCNNPS 583  
 534 ldgtemcapghcfkghciwltptd-llkrtdgswagwspfgscstrcglvqkfrtrqcdph 592  
 584 PAYGRCCLPMEYQVNCSEECRGTEDEPRACAKRNYUYVHONAKSHVYEEPDDA 643  
 593 panggrtcsqlydfiglscrqdcdsladtreecrqwdlytengdaqhncipheindak 652  
 644 QKCELICOSADTGDUVFMNQVNHDRCSYRDPSYVCARGECVPGCDKEVSGMKADKC 703  
 653 erchlycestrtgevsmkrmvhgtrcsykdaflcvrgdcrkvgdgylgssqgdkc 712  
 704 GYCGGDNSSHCKTVKGTGKASKAGALKVQIPRAGAHQIDLEAKSPHNSYVKNQVYGS 763  
 713 gycggdnshcktvvkgftrtspknhgylkmtfepagahlllqvedadshlnlawnletgk 772  
 764 FTLMPKCK-EATSTPTAMGLEWEDAVEADAKESLKTSGPPEALAIALLPTEGSPRSL 822  
 773 fltlnendvdaaskftlamgyewyredgedretqltmgprlghltvlvip--vgdtrvsl 830  
 823 AKYVYHEDLLPLGNNVLEEMDTYEMALKSWAPCSKACGGGIGTOTTKGCRRRRHHM 882  
 831 tykymihedsin-vddnnlleedsvyewalkkwpsckpogsgsqtllkygctrrldhkm 889  
 883 VQRHLCCHKRRPKPLRRNCQNHPSOPVWVWEWAGSACSGKGYVOTRGIOCLLPISNG 942  
 890 vhrfgcaalskpkairracrpqesqrvwlvgeewescqtrtgmgyvtrcqlphdn 949  
 943 THKVPARACAGDRPEARPCILRVPCRAMRLGAMSOCSATCGEGIGQROVVCFRTNANSL 1002  
 950 ttrshahncdarpesttrascrtelcpgtrvragwsgsvtvcgqgtgerplcrtadstf 1009  
 1003 GHCBGDRPDYQVCSLPAACGGNHQNSTYRADVWELGTPEGWVQSGPLHINKISSMCA 1062

DB 1010 glgceetpetarctrlgpcprnlstdpskksyv-----vqwlstrpdpdsprtkiss--- 1059  
 QY 1063 AEPCTGDRSVFCOMEVLDKRCSTPGYHRLCCVSC-----IKKASGNPDP---DPG 1110  
 DB 1060 kphcgdkstfcrmevstryslpygnyklsckscnlynnltnvegrlepppykhdidvtf 1119  
 QY 1111 PTLSPRSTRGSPPLPGQPDADAAEPPGKPRFGSEDHQHGKATQLPGALDTSRFGQHP-F 1169  
 DB 1120 mptlprvtwemvtrpsptlrvleplnassnatcdhpetnavdepykklngldevgprnl 1179  
 QY 1170 APETPIP 1176  
 DB 1180 iprrpsp 1186

RESULT 6  
 ABG22373  
 ID ABG22373 standard; Protein; 1328 AA.  
 AC ABG22373;  
 DT 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #22364.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 PE 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PA (HYSE-) HYSEQ INC.  
 DR Dmanac RT, Liu C, Tang YT;  
 DR N-PSDB; AAS86560.  
 WPI: 2001-639362/73.  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 20; SEQ ID NO 52732; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques for restoring normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on human amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed



CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1328 AA;

Query Match 47.2%; Score 3279.5; DB 22; Length 1328;  
 Best Local Similarity 53.1%; Pred. No. 1.2e-224;  
 Matches 630; Conservative 153; Mismatches 349; Indels 55; Gaps 20;

```

QY 16 CALCTAAGSR--TPELHLSGL---SDYGVTPCSTDFRGNFLSHVYSGPAASAGSMV 70
D 40 crrlllaasrlhqeprgevggprlgbaerllayprtdagqlvshvsa-atsagrar 98
D 71 DPPFLPRHSHLRVARSPLHGGTGLMGRGRHSITFNATVFEKELHLRLRPRLVLP 130
D 99 raap-----vtrpsfpggn--eepgsh-lfynvvtigrdhlrlrpnarlvp 144
QY 131 GSSVEMQDFRELEFRQPLROECVYTGVTGMPGA-AVAISNCODLACLRDSTDFIEP 189
D 145 galmevgekgqltrvepllgsclyvgdvaglaeassvalencdglaglimeeeffiep 204
QY 190 LERGOQKEA--SGRHHVYRREAQOEAEPDQ--DLHNEAFGLGDLPNLLGLVQDQDGT 247
D 205 lckglaaeaeagrvhvyrrptsppllggpdaltdgaslslsralgyleehans 264
QY 248 ERK--RRNAKPGSYSEVLLVVDSDVFRHGEKHQYQNVLTLMNVDEIYHDESLGYHIN 306
D 265 rrrarhaaddynlevllgydvsvfngkhenyqkylclmltveiydeslghahlv 324
QY 307 ALVRLIMGYKQSLIERGNPSRLSEQVCRMAHSQQRDPSSAENHNVFETRDQEP 366
D 325 vivrllllygksmllelgnpsqslenvcrwaylqgkptgldexyhdhaifltqdfg 384
QY 367 S---GYAPVTGMCPLRSCALNHEDGSSAFVIAHENGHGLMEHQGNGCADETSLS 423
D 385 sgmgypapvymchprvscclnhdgfsatvvaheghvlgmehdgqgrcdevrlls 444
QY 424 VMAPLVQAARHFRHWSRCSKLESLRYLPVYDCLLDDPFDPANQPDELGINVSMDEQR 483
D 445 imaplvgaaafnrflmwsrscgelsrylshydcilddpfaidwpaipqjlylshmeqcr 504
QY 484 FDPGSGYQTCIARFTRPCKQIMCSHPDNPFEKTKKGPRLDTECAPKWCFCGHIWK 543
D 505 fdrlgyymcaltittdpcqkwlscnmdpylcktkppldgtmcapqkchckfncial 564
QY 544 SPEDOTYGDGGSWMTKFGSGSRSCGAGVRSRSCNNPSPAYGGRPCGPMFEYQVCS 603
D 565 tpd-llkrdsgwagwspfgscstcgtlyvklftrgcdmhpnangrtcsqglaydtqlcsr 623
QY 604 EECGTIEDFRAOQCAKRSYVYHONAKHSWYPERDDAQKCELLCQASDTGADVFMNO 663
D 624 qdcpsdlaadireeqgrwdlyfehgdaghlwlphehrdakechlyceeregeevsmkr 683
QY 664 VHDGTGCSYRDYSYCARGECPYVCGDKVEGSMKADDCGVCGGDNHCHRTVKCTGLKA 723
D 684 mwhdgtrcsykdaflscvqgdcvkvgcdgvlsgskedcvcvggdshckvkvkfltrs 743
QY 724 SKQAGALKIQLIPAGARHIQIEALEKSPHSRYKNQVTSFLINKPKG--EATSRTFTMG 782
D 744 pkhgyikmtefpagrarhlligevatshlavklegkflleendvdasskflfiamg 803
QY 783 LEWEDAVEDEAKESLTSGLPFAIALIALPTEGGPRSSLAYKYVHEDLPLIGSNNVL 842
D 804 veweyrdeqrelclgmplhgtlvtlvip--vgdtrvslcykylhbedslh-vddnvl 860
QY 843 LEEMDTYEMALKSWAPCSKACGGAGIOFTYKGRRRDHMMQRHLCDBKKRPTRRCN 902
D 861 eedsvvyewalkkwspcsrpgqsgqfktlygtrrllldhmvmvrtgcaalskpkalraen 920
QY 903 QHPCSQPVWTEEMACSSRSGKLGQVTRGICQLPLNSGTHKVPAPAKAGDPEARRP 962
D 921 pgecsqpvwvtegewepscgtcgrtgmyrvsvrciqlndnttrshahcndaapestra 980

```

```

QY 963 CURVPCPAOWRLGAMSOCATGEGEIQOQRYVCRNMANSLGCEGDRDPVQYCSIPACG 1022
D 981 cserlcpgrwtagpwsqsvlcognlgerpvcrtadslgylqeeepelartrcrlgpcp 1040
QY 1023 GHNQSTVRADYVWELGTREGQWVPOSGPLHPINKISSMCAEPCTGDSRYFCOMEVLDYR 1082
D 1041 rnsdpskksy-----vqlsrpdpdsplrlkss---kghcgqskslfctmervlsry 1090
QY 1083 GSIPGYHRLCCVSC-----IKKASGNRGP----DPGETSLPFTSGSPLPGRQDP 1130
D 1091 cslggkrltlqkscllyunpntvegrleppqkndldvfmpltpvtvamevrpsstp 1150
QY 1131 ADAAEPKPGTCSGSDHOGHGRATOLPGALDTSSTPGQHP-PAPETPRP 1176
D 1151 levpnasslnatedhpnetnavdepyklhngldevqpnllprrpsp 1197

```

# RESULT 7

AAW47030 standard; Protein: 1206 AA.

AAW47030;

06-JUL-1998 (first entry)

Bovine N-proteinase.

N-proteinase; cattle; collagen; antibody; rheumatoid arthritis;

fibrosis; Ehlers-Danlos disease; diagnosis; therapy.

Homo sapiens.

```

FH Key Location/Qualifiers
FT MISC-difference 142 /note="encoded by CCC"
FT MISC-difference 207 /note="encoded by CCC"
FT MISC-difference 464 /note="encoded by CCG"
FT MISC-difference 545 /note="encoded by CCC"
FT MISC-difference 549 /note="encoded by ATC"
FT MISC-difference 810 /note="encoded by CCG"
FT MISC-difference 980 /note="encoded by ATC"
FT MISC-difference /note="no equivalent codon in nucleotide sequence"

```

WO9800555-A1.

06-JAN-1998.

03-JUL-1997; 97WO-US12427.

02-JUL-1997; 97US-0886333.

03-JUL-1996; 96US-0021203.

(COLI/) COLIGE A.  
 (LAP/) LAPIER C.  
 (PROC/) PROCKOP C. J.

Collage A, Lapierre C, Prockop DJ.

WPI, 1998-086980/08.

N-PSDB; AAV06594.

Polynucleotide sequence encoding human N-proteinase - used to produce mature collagen in vitro and antibodies to treat fibrosis and rheumatoid arthritis

Disclosure: Fig 5; 49pp; English.

This polypeptide comprises human N-proteinase, the enzyme responsible for cleaving N-propeptide from procollagen to produce mature collagen. The amino acid sequence was deduced from a cDNA sequence (see AAV06594) derived from bovine cDNA clones. Bovine N-proteinases can be produced using a claimed method in which a host cell is transformed or transfected with an N-proteinase cDNA, and the N-proteinase is isolated from the medium. The N-proteinase can be used for the production of mature collagen in vitro and for the production of antibodies which may be used for diagnosis and therapy of diseases including fibrosis and rheumatoid arthritis. The N-proteinase may also be administered to treat a disease resulting from insufficient production of N-proteinase, such as Ehlers-Danlos disease. Human N-proteinases (see AAM47028-29) are also disclosed.

Sequence 1206 AA:

Query Match 47.0%; Score 3270.5; DB 19; Length 1206;  
 Local Similarity 51.7%; Pred. No. 4.4e-224;  
 Matches 636; Conservative 160; Mismatches 356; Indels 79; Gaps 21:

```

7  LLSTLPLHCACTAAGSRPELHLSGLSDYGVTPCSTDFGRFELSHVSGPAAASAG 66
17  LLLLLLPadarlaaaadpppgpghgeerlavrvtdaggrlvshvsa-atapag 75
67  SMVVDPEPTLPRHSHLRVARSPLHGGTLMGVRHSLYEVNVAFGKELHLRLPNR 126
76  ---Vrrtraapagpgls-ggseedpgg-----rlfynvtvgrlhlrlpnar 121
127  LVVGGSSVENQEDRELFRQPLROECVYTGVTGM-PGAVAISNCGGLRLRTSDTF 185
122  lvapgatvewgsegatrvellgtcltygdvaglaessvalscnccagllrlmeeef 181
186  FIEPLERGOQKEA-SGRTHVYVYRRAVVOQEMAPDG-----DLHNAFGLGDLPNLGL 239
182  fliepleglaakeaggrvrvhvnrtftr--ppplgspaldgtisadsidsralgy 239
240  VGDLGDTERR-KRHAKEGSIIEVLYVNDVYRFHGEKENVQVYVITLINIYDEIHYDE 298
240  leevnsarrmrhaaddnyllgyddsvvqfngtehvqgyltllmilyelxhde 299
299  SLGVHINATLRLVWGRQSLIERGNPSRLQVCRMAHSQORODPSHAHHHNVF 358
300  slgahinvvrllyllygkmslleiagnpsqlenvcrcway-lgkprdtldheydhalf 359
359  LTRQDFGS---GYAPVTGMCHPLRSCALNHEDGSSAFVLAHETGHVLGHEHNGONGC 415
360  ltrqdfgsgmggyarpyvcmchpvrscclnhedgfssafvahalghvlgmehdggnrc 419
416  ADETSLSGVMAPIVOAAFRFPHMSKSLRLSPYDCLLDPPFPAMPQPELPGIN 475
420  gdevrlgsmaplvaafhrfwrscsqelsrlshysdclrdffcthdwpalpqlp9lh 479
476  YSMDOCRDFEGSYOTCLARFTEPCOLWCSPHDNPFCKTKKGPPLDTECAPGKWC 535
480  ysmecgrdfglgymmtcafrtfqpcxqlwcpshpndpyfcktkkppldgtumcapkhc 539
536  FKGRHITKSPEDTYGDDGWSWTKESGSRSCGGVRSRSCONPSPAPAGRPCLGPM 595
540  fkghtwtlftd-llkrdgnwagwspfgscrtcgvtgvtfrtqcanpnapaggtcsgla 598
596  FEYQYNSSECPGTIEDPRAQCAKRNSTYHQNAKHSWVPEPDDAQKCELICQASDT 655
599  ydfqgncsdcpdaladfrteegctqwdlyflehgdqhlwphrhdakerchlycesket 658
656  GDUVENQVYHDGTRCSTRDPYVVCARCECVPGCDKEVGSMAKADKCGVGGDNHSHRT 715
659  gevvsmlkrmhvgtcrtsykdafslcvrgdckvsgcdgylysgskgdcgvsggdshshk 718
716  VKGTLTAKAKOGALKIVDIPAGARHIQIEALEKSPHRSVYKNOVTGFTILNPKCK-DAT 774

```

```

Db 719  vkgtfrspkklgylkmtelpagarhlllgadtstshlavknuletykllneendvdpn 778
QY 775  SRPTFAMGLEDEDAVEDAKESLKTSGPLPEALIALPPTREGRRSLAKKYIHEDLP 834
Db 779  sktflamqveweyredqgrellqtmprhgtltvlp--egdarisltykymlnedsln 836
QY 835  LIGSNVNLLEMDYEMALKSMAPCSKACGGIOTFYGGCRRRRDHMMVOHILCDHKKRP 894
Db 837  -vdannvledsavyewalkkwpsscpkpcgsgqfkygcrtridhkmvnhgfdcsvskp 895
QY 895  KPIRRRCNOHPCSPVWVTEEMGACSRSGKLVQTRIGQCLPLNSGTIKVMARKACAG 954
Db 896  kafrtcmqgecsqgvwtgsewpcscgrltmqvsvrvcvpllnhtlsvtkhncd 955
QY 955  DRPEARRPCLVPCPAQRLGAMS-OCSATGCEGIGQROVQCRFMANSLGHCEDRDPDV 1013
Db 956  arpegtracnrelcpgrwraswagcsvtgngntdgetrpylctradsfygcreepeta 1015
QY 1014  QVCSLPACGNGHONSTVRADYWELOTPBGQVWPQSGPLHPINKISSMCAEPTCGDSVF 1073
Db 1016  ricrlgpcrntsdpskksy-----vqlsrpdpnspvqetss--kgrcgdksvf 1065
QY 1074  COMEVLDRKCSIPGTHRLCCVSC-----IKKASGPNPGPPGPTSLPFGSPGSL 1124
Db 1066  crmewlsrscslpygnklcscnphnlltdvddraeppskhndieclmptlsvlytm 1125
QY 1125  ---PGROPADAAEPKPGTGESEDHQHRANQLPGALDTSRGTQHPAPETPIPGASWS 1181
Db 1126  ewgpppglpvleprltsstnatedhpefnavdpykl-----pgledevqpnllp----- 1176
QY 1182  ISPTTGGLPWCGWTPPTPVPEDKQPGEDL 1212
Db 1177  -----rrpspyektlnqrigel 1193

RESULT 8
AAV53900
ID AAV53900 standard; Protein; 969 AA.
XX
AC AAV53900;
XX
DT 13-MAR-2000 (first entry)
XX
DE Amino acid sequence of a bovine PINP protein.
XX
KW GON-1; metalloproteinase; cell migration; modulator; PINP;
KW metalloproteinase domain; thrombospondin domain; abnormal cell migration;
KW organ shaping; sterility; cancer metastasis.
XX
OS Bos sp.
XX
PN WO961656-A2.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11918.
XX
PR 29-MAY-1998; 98US-0087170.
PR 13-APR-1999; 99US-0129023.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Kimble JE, Bieleloch RH;
XX
DR WPI; 2000-072633/06.
XX
PT Identifying modulators of proteins containing metalloproteinase and
XX thrombospondin domains, potentially useful for controlling cell
XX migration and organ shaping
XX
PS Disclosure; Fig 1C; 60pp; English.
XX
CC The present sequence represents a bovine PINP protein. PINP is

```

CC a metalloproteinase. The specification describes another related  
 CC metalloproteinase, a C. elegans GON-1 protein, that lacks a transmembrane  
 CC domain and possesses a predicted metalloproteinase domain between residues  
 CC 269-456. In C. elegans hermaphrodites, GON-1 is required for migration of  
 CC two distal tip cells to produce elongated tubes, whereas in males, GON-1  
 CC is required for migration of a single linker cell to produce a single  
 CC elongated tube. The protein is used in the method of the invention. The  
 CC specification describes a method for identifying a modulator of a  
 CC protein that contains a metalloproteinase domain and a thrombospondin  
 CC domain. The method comprises treating a target organism, having a  
 CC developing gonadal cell that is responsive to the protein, with a test  
 CC compound, and determining any change in migration or shape of the cell  
 CC attributable to the test compound. The compounds identified are  
 CC potential therapeutic modulators of abnormal cell migration and organ  
 CC shaping, e.g. for rendering animals (specifically nematodes) sterile  
 CC and for inhibiting cancer metastases.

Sequence 969 AA:

Query Match 42.6%; Score 2962.5; DB 21; Length 969;  
 Best Local Similarity 57.4%; Pred. No. 2,9e+202;  
 Matches 559; Conservative 123; Mismatches 259; Indels 33; Gaps 14;

QY 7 LLSYLLPICALCTAAGSRTPETLHLSGLSDYGVTPCSTDFRGRFLSHVSGPAAASAG 66  
 DB 17 LLLLLLPpadarlaaaadppgpgbgaerilaavprtdagrlvshvsa-atapag 75  
 QY 67 SMVVDPTPLPRHSHLRVARSPLHGGTLMGCRGRISLRYNVVPEKELHLRLRNR 126  
 DB 76 ---vrrraapagipgls-gseedpgg-----rlllynvvtfgdrhlrlrpar 121  
 QY 127 LVVPGSSVMEQEDFRLRQPLROECVYTGVTGM-PCAATAISNCDCLAGIRFSDTF 185  
 DB 122 lvpagatvewqesgslrvepllgcllygdvaglaessvalncdglalimeeef 181  
 QY 186 FLEPLERGOOEKKA-SGRHVVYRKREAVOQEAEPDG-----DLHNEAFGLDPLNLGL 239  
 DB 182 fleplekylaaakeagqrvhvhyhrltsr--pppllpgqaldtqisadsidsralg 239  
 QY 240 VGDOLGDTERRK-RRIAKPESYSIEVLVVDSSVYRFGKEHVMONVLTIMVETIYHDE 298  
 DB 240 leevnssrrmrtaadddynlellygvdsvqfthetvqkyltlmliveiyhde 299  
 QY 299 SLGVHINLALVLIWGYRQSLIERGNPSRLSEQVCRMAHSQRODPSSHAEHHDHVF 358  
 DB 300 slgahlnvrlilllsyqksmllelgnpsqlenverwaylqkprdtchdeyhnaif 359  
 QY 359 LTRDFGFS---GYAPVGMCHPLRSCALNHEDGFSSAFVLAHTGYLVGMEHDGNGC 415  
 DB 360 ltrqdfgspmggyavpvgmchprvsccllnhedgfssafvalhetghvlgmehdgngnc 419  
 QY 416 ADETLGSMAPLVOAFAHFRHMSKSLERLRLPSYDCLDDPFPAMPQPELPGIN 475  
 DB 420 gdevlrglmaplvgaefhrfhwercsqelsrlylshydcldrddpfhndwpaipqjln 479  
 QY 476 YSMDOGRFDFGSGQOTLARTRPECKQIWCSPNDNPFCKTKKGPDLGTGCAPKWC 535  
 DB 480 ysmdegctidfglymmctalrtdpcqlweshpndnpycklkkppldgtmcapgknc 539  
 QY 536 FKSHCINRSPQOTGODGGSMTKFGSCSGGGRSSRCNNNSPAYGGRPLCLGPM 595  
 DB 540 fkgphwtltpd-llkrqngwawspfgscstcgvykkltrtdcpdpbpanagrtcsqila 598  
 QY 596 FEYOVNNEECPGTIEDFRAOQCAKRNSSYYVHONAKHSWPPEPDDAOKCELTICOSAPT 655  
 DB 599 ydfqfscsqdcdpdaadfreesqcrwldylehngdahqhwlphehtrakechlyesket 658  
 QY 656 GDVVMONVVDGTRGSRDPYSYVARGCEVPVGCCKEVSMAADCKCGCGGSDNSHCKT 715  
 DB 659 gevsmskmvndgtcsykdaflscvrgdcrkyvgcdvlgsskqekcvgcgdsnshckv 718  
 QY 716 VKGTGLKASQAGALKLVQIPAGARHIOLEALEKSPHRSVYKQVGVGFIANKGR-EAT 774

DB 719 vkgtfsspkklgyikmfepagarlhlqeadtshlavlknlegrkllneendvgn 778  
 QY 775 SRPTTAGLMEBAVEDAKESLTKSGPLPEALILALPTREGPSRLAKYVTHEDLLP 834  
 DB 779 sktliangvewyrdedgretlqtmplhgtlvlp--eddarlslykymhdedln 836  
 QY 835 LIGSNVLLSEMDTYEMALKSMAPCSKACGGIOFTKYGCRRRRDHMYORLCHKKRP 894  
 DB 837 -vddnvlveddsvyewalkkwspsckpggsgqftkygcrirrlhkmwhrgfcdsvskp 895  
 QY 895 KPIRRRCNORPCOPVWVIEWMCACRSGCKIGVOTRIGIQLPLSNCTHKYMPAKACAG 954  
 DB 896 kairrtcnpgcsqpvwlvgewpcscrgtgmgyrsvrcvplhnttsvthkncnd 955  
 QY 955 DRPEARPCLRVPC 968  
 DB 956 arpegrracnrelc 969

RESULT 9  
 AAM47029  
 ID AAM47029 standard; Protein: 566 AA.  
 XX  
 AC AAM47029;  
 XX  
 DT 06-JUL-1998 (first entry)  
 XX  
 DE Human N-proteinase (70 kDa short form).  
 XX  
 KW N-proteinase; human; collagen; antibody; rheumatoid arthritis;  
 KW fibrosis; Ehlers-Danlos disease; diagnosis; therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09800555-A1.  
 XX  
 PD 08-JAN-1998.  
 XX  
 PF 03-JUL-1997; 97MO-US12427.  
 XX  
 PR 02-JUL-1997; 97US-0886333.  
 PR 03-JUL-1996; 96US-0021203.  
 XX  
 PA (COL1A) COL1GE A.  
 PA (LAP1) LAP1ERE C.  
 PA (PROC) PROCKOP D J.  
 XX  
 PI Collige A, Lapiere C, Prockop DJ;  
 DR WPI: 1998-086980/08.  
 DR N-PSDB: AAV06593.  
 XX  
 PT Polynucleotide sequence encoding human N-proteinase - used to  
 PT produce mature collagen in vitro and antibodies to treat fibrosis  
 PT and rheumatoid arthritis  
 PS Disclosure: Fig 2B; 49pp; English.  
 CC  
 CC This polypeptide comprises the 70 kDa short form of human  
 CC N-proteinase, the enzyme responsible for cleaving N-propeptide from  
 CC procollagen to produce mature collagen. The amino acid sequence  
 CC was deduced from a cDNA sequence (see AAV06593) derived from human  
 CC skin fibroblast cDNA clones. The 130 kDa long form (see AAM47028) of  
 CC human N-proteinase was also identified. Human N-proteinases  
 CC can be produced using a claimed method in which a host cell is  
 CC transformed or transfected with an N-proteinase polynucleotide  
 CC sequence, cultured in an appropriate culture medium, and the  
 CC N-proteinase is isolated from the medium. The N-proteinase can be  
 CC used for the production of mature collagen in vitro and for the  
 CC production of antibodies which may be used for diagnosis and  
 CC therapy of diseases including fibrosis and rheumatoid arthritis.  
 CC The N-proteinase may also be administered to treat a disease



Db 659 asrdicngicKnygcdfeidsamedrcgcyhngstchtvsgtfeea eqlyvwdvgl 717  
QY 735 IPAGARHIOIFALEKSPHSRVKNQVTGSLFNLKGEKASRTFTAMGLEMEDAVEADAKE 794  
Db 718 ipagarelrigeveaanflalrisedpekyflngwrlgndyvaqgtlftyarrgnwe 777  
QY 795 SLKTSGRPLPEAIALALPRTGGRSSLAKYVHEDLPLIGSNNVLEEMDTYEMALK 854  
Db 778 nltsprpkpewlqll-fgesnp--gvhyeylthre---agshdev--pprvfshyg 828  
QY 855 SWAPCSACGGGIGFTYKGGRRRHHMYORHLCDHKRRPKPIRRCONHPCSQPV-WWT 913  
Db 829 pwtktvlgcygrygnyvclerqagrvdeeh-cdplgprddqgrkceqpc--parwa 885  
QY 914 EEWGACSRSCCKLGVOIRGICQLPLPLSNGTHKVMAPAKACAG-DRPEARRRCLR-VPCPAQ 971  
Db 886 gewqlcsscgpglstrravlcitrsvgldegsaleppacehlprppteletcmrhvpcpat 945  
QY 972 WRIGAMSQCSATCGEGIOQRQVVCRTNANSLGHCEGDRPDVTQVCSLPACGGNHQNSTYR 1031  
Db 946 wavgwsgcsvtcgegtqrtnvlic-tndtyvpcdeaqdpasevtcsiplcr----- 995  
QY 1032 ADVWELCT--PEGQWVPOSG-----PLHPINKISSMCAEPCTGDRSVFCQMEYL 1079  
Db 996 ---wplgltlqpegsgsgssshelIneadfiphlprpspasspkpqtmgnaleeaepel 1052  
QY 1080 DRYCSIPG-----YHRLCCVSCIKKAS-GPNRPGD-----PGTSLPFFSTPG 1121  
Db 1053 d-----lpgvfvddfydydnfInfhedlisygspeepdlagtgdrtpphshpaapstg 1108  
QY 1122 SPLPGQDPADAAE---PRGKPTGSEDHOGH---ATOLPG-ALDTSSPGTQHPF-AP 1171  
Db 1109 spvpateppaakegylgwpsps-pwpsqagrspppsseqtpgnplInflpeedtpigap 1167  
QY 1172 ETPPLPGASWS-----ISPTTP-----GCLPMGWTQTPRPVPEDKGQGED 1211  
Db 1168 dlglpelswprvstdglqtpatpbesqndfpygkdsqslppwrtdrtnevfkdeeep--- 1224  
QY 1212 LRHPGTSLPADLPGRPPE---PCHPTGT 1236  
Db 1225 ---kgrgap-hlprpsstlpplspvgs 1248

Fetch completed: August 28, 2002, 20:10:28  
Time: 1100 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 22:35:24 ; Search time 61.11 Seconds

(without alignments)

361.697 Million cell updates/sec

Title: US-09-938-330-22\_COPY\_1040\_1249

Perfect score: 1206

Sequence: 1 PEGQWVPGSGPLHPINKISS.....PFGTFLCVLRDSQLRGHT 210

Indexing table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Database: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID55/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID55/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID55/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020	84.6	1048	22	AA85695*
2	217.5	18.0	1311	19	AAW47028*
3	211.5	17.5	1203	22	AA850004*
4	211.5	17.5	1206	19	AAW47030*
5	204	16.9	1201	21	AA821254*
6	204	16.9	1205	22	AA873550*
7	188.5	15.6	1328	22	AA823373*
8	175	14.5	762	16	AA880327
9	175	14.5	762	16	AA880324
10	175	14.5	762	16	AAW49713
11	175	14.5	762	19	AAW49715

12	175	14.5	762	19	AAW57666	Collagen-like poly
13	175	14.5	762	19	AAW57668	Collagen-like poly
14	175	14.5	762	19	AAW53535	Amino acid sequenc
15	159.5	13.2	635	22	AA82860	C glutamicum prote
16	158	13.1	682	16	AA80330	protein polymer ad
17	158	13.1	682	19	AAW49717	Collagen-like poly
18	158	13.1	682	19	AAW57670	Collagen-like poly
19	154	12.8	417	18	AAW26354	Collagen-like poly
20	154	12.8	417	18	AAW26354	Collagen-like poly
21	154	12.8	417	21	AAW53531	Amino acid sequenc
22	154	12.8	417	21	AAW78291	CLP 3.7 amino acid
23	154	12.8	837	16	AA80320	Collagen-like prot
24	154	12.8	837	19	AAW26353	Collagen-like prot
25	154	12.8	837	19	AAW57659	Collagen-like poly
26	154	12.8	837	21	AAW53530	Amino acid sequenc
27	154	12.8	837	21	AAW78290	CLP 3.7 amino acid
28	152.5	12.6	829	16	AA80334	Protein polymer ad
29	152.5	12.6	829	19	AAW57673	Collagen-like poly
30	152.5	12.6	829	19	AAW57673	Collagen-like poly
31	151	12.5	448	21	AAW53201	Macaca mulatta rha
32	150	12.4	330	19	AAW57645	Collagen-like poly
33	150	12.4	408	22	AAW07539	Repetitive protein
34	150	12.4	408	22	AAW72737	CLP protein sequen
35	150	12.4	408	22	AAW64007	Human atrophin-1 r
36	149.5	12.4	408	22	AAW7406	Type III procollag
37	149.5	12.4	1012	20	AAW17406	Mouse Rho target p
38	148	12.3	1664	19	AAW3106	Pro-alpha (I) (I)
39	145.5	12.1	1255	19	AAW52249	Truncated pro-alpha
40	145.5	12.1	1255	20	AAW6733	Drosophila melanog
41	145	12.0	623	18	AAW12843	Collagen type III
42	145	12.0	626	18	AAW12842	Novel human diagno
43	145	12.0	1180	22	ABW5175	
44	145	12.0	1466	22	ABW50291	
45	145	12.0	1469	22	ABG15191	

#### ALIGNMENTS

RESULT 1	AA85695 standard; Protein: 1048 AA.
ID	AA85695:
XX	AA85695:
AC	29-OCT-2001 (first entry)
DT	Larvax carcinoma associated protein-1 (LarCAP-1) polypeptide.
XX	Larvax carcinoma associated protein-1; LarCAP-1; carcinoma; cytostatic;
XX	antimetastatic; antiarthritic; osteopathic; immunosuppressive; human;
KW	immunostimulant; cerebroprotective; vasotropic; antiinflammatory;
KW	dermatological; cardiact; vaccine.
XX	Homo sapiens.
OS	WO200159133-A1.
PN	16-AUG-2001.
XX	12-FEB-2001; 2001WO-EP01525.
PE	14-FEB-2000; 2000EP-0102955.
XX	(MERE ) MERCK PATENT GMBH.
PR	Duecker K, Hentsch B, Hoheisel J, Frohme M;
PA	WPI: 2001-483569/52.
XX	N-PSDB; AAH47048.
DR	Larvax carcinoma associated protein-1 polypeptide for the treatment of
XX	carcinomas, metastasis, arthritis, osteoporosis, immune disorders,
PT	



```

PR 22-DEC-1999; 99US-0171503.
PR 22-FEB-2000; 2000US-0183792.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (SMIK-) SMITHKLINE BEECHAM CORP.
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
PA (JONA/) JONAK 2 L.
PA (TRUL/) TRULLI S H.
PA (FORN/) FORNWALD J A.
PA (TERR/) TERRETT J A.
XX
IRUELA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
Fornwald JA, Terrett JA;
WPI: 2001-025136/03.
XX
PT METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis -
XX
PS Disclosure: Fig 3; 768bp; English.
XX
CC The present invention relates to human METH1 and METH2, (ME for
CC metalloproteinase and TH for thrombospondin; see AAB50002 and AAB50003).
CC The present sequence is bovine metalloproteinase, which was used in a
CC sequence homology comparison with the METH proteins. METH can be used
CC for inhibiting angiogenesis in an individual, and for treating cancer,
CC benign tumors, an ocular angiogenic disease, rheumatoid arthritis,
CC psoriasis, delayed wound healing, endometriosis, vasculogenesis,
CC granulations, hypertrophic scars, nonunion fractures, scleroderma,
CC trachoma, vascular adhesions, myocardial angiogenesis, coronary
CC collaterals, cerebral laterals, arteriovenous malformations, ischemic
CC limb angiogenesis, Osler-Webber syndrome, plaque neovascularisation,
CC telangiectasia, haemophilic joints, angiodiroma, fibromuscular
CC dysplasia, wound granulation, Crohn's disease or atherosclerosis. METH
CC can also be used in birth control. METH can also be used in diagnostic
CC methods for the prognosis of cancer.
XX
SQ Sequence 1203 AA;

```

```

Query Match 17.5%; Score 211.5; DB 22; Length 1203;
Best Local Similarity 29.1%; Pred. No. 6e-07;
Matches 52; Conservative 24; Mismatches 68; Indels 35; Gaps 4;
4 OMVPOGSLHPINKISTTEPCGDSRVCOMEVLDRCISIPGYHRLCCVSC----- 54
Db 1035 gwlsrpdnspvqelskgrcgqskvrcmevlsrscysipgynklcckscnphnltlv 1094
QY 55 IKKASGPNPGDPGPTSLPPFSTPGSPPL---PGQDPADAAEPKGRPGSDHOGHGRATQ 111
Db 1095 ddraeppsgkhndieemltvptlwmvqppgipilevplntsstnatdhpethnavd 1154
QY 112 LUGALDTSFPGQHPAPETPIGASWSISPTTGGLPWGTQTPTPYVEDKSGPGEDL 170
Db 1155 vpykl---pgjledevqpnllp-----rrpspyektrnrgl 1190

```

```

RESULT 4
AAM47030
ID AAM47030 standard; Protein; 1206 AA.
XX
AC AAM47030;
XX
DT 06-JUL-1998 (first entry)
XX
DE Bovine N-proteinase.
XX
KW N-proteinase; cattle; collagen; antibody; rheumatoid arthritis;
fibrosis; Ehlers-Danlos disease; diagnosis; therapy.

```

```

XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 142 /note="encoded by CCC"
FT Misc-difference 207 /note="encoded by CCG"
FT Misc-difference 207 /note="encoded by CCG"
FT Misc-difference 464 /note="encoded by CCC"
FT Misc-difference 545 /note="encoded by ATC"
FT Misc-difference 549 /note="encoded by CCG"
FT Misc-difference 810 /note="encoded by ATC"
FT Misc-difference 980 /note="no equivalent codon in nucleotide sequence"
XX
XX W09800555-A1.
XX
XX 08-JAN-1998.
XX
XX 03-JUL-1997; 97WO-US12427.
XX
XX 02-JUL-1997; 97US-0886333.
XX 03-JUL-1996; 96US-0021203.
XX
XX (COLL/) COLIGE A.
XX (LAPI/) LAPIERE C.
XX (PROC/) PROCKOP D J.
XX
XX Collige A, Lapiere C, Prockop DJ;
XX
XX WPI: 1998-086980/08.
XX N-PSDB; AAV06594.
XX
XX polynucleotide sequence encoding human N-proteinase - used to
XX produce mature collagen in vitro and antibodies to treat fibrosis
XX and Rheumatoid arthritis
XX
XX Disclosure: Fig 5; 49pp; English.
XX
XX This polypeptide comprises human N-proteinase, the enzyme
XX responsible for cleaving N-propeptide from procollagen to produce
XX mature collagen. The amino acid sequence was deduced from a cDNA
XX sequence (see AAV06594) derived from bovine cDNA clones. Bovine
XX N-proteinases can be produced using a claimed method in which a
XX host cell is transformed or transfected with an N-proteinase
XX cDNA polynucleotide sequence, cultured in an appropriate culture medium,
XX and the N-proteinase is isolated from the medium. The N-proteinase
XX can be used for the production of mature collagen in vitro and for
XX the production of antibodies which may be used for diagnosis and
XX therapy of diseases including fibrosis and rheumatoid arthritis.
XX The N-proteinase may also be administered to treat a disease
XX resulting from insufficient production of N-proteinase, such as
XX Ehlers-Danlos disease. Human N-proteinases (see AAM47028-29) are
XX also disclosed.
XX
SQ Sequence 1206 AA;

```

```

Query Match 17.5%; Score 211.5; DB 19; Length 1206;
Best Local Similarity 29.1%; Pred. No. 6e-07;
Matches 52; Conservative 24; Mismatches 68; Indels 35; Gaps 4;
4 OMVPOGSLHPINKISTTEPCGDSRVCOMEVLDRCISIPGYHRLCCVSC----- 54
Db 1038 gwlsrpdnspvqelskgrcgqskvrcmevlsrscysipgynklcckscnphnltlv 1097
QY 55 IKKASGPNPGDPGPTSLPPFSTPGSPPL---PGQDPADAAEPKGRPGSDHOGHGRATQ 111
Db 1098 ddraeppsgkhndieemltvptlwmvqppgipilevplntsstnatdhpethnavd 1157

```



RESULT	8
ID	AAR80327 standard; Protein: 762 AA.
XX	
AC	AAR80327;
DT	18-APR-1996 (first entry)
XX	
DE	Protein polymeric adhesion substrate 1-B.
KW	Pendent group; repeating unit; enzyme recognition site; sealant; fibrin;
KW	enzymatic cross-linking; biocompatible material; structural integrity;
KW	medical adhesive; wound closure; tissue repair; transglutaminase.
OS	Synthetic.
XX	
PN	WO9523611-A1.
XX	
PD	08-SEP-1995.
XX	
PF	03-MAR-1995; 95WO-US02728.
XX	
PR	03-MAR-1994; 94US-0205518.
XX	
PA	(PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX	
PI	Cappello J;
XX	
DR	WPI, 1995-320413/41.
PT	Protein polymers comprising repeating units and sequences - capable
PT	of enzyme-catalysed covalent bond formation useful as a
PT	biocompatible material for wound closure and tissue repair
PS	Example 4; Page 45; 138pp; English.
XX	
CC	The amino acid sequence of the protein polymeric adhesion substrate
CC	(PPS) 1-B. The protein comprises 10 repeats of the PPS1-B monomeric
CC	repeat (AAR80326) which consists of the CLP 3.7 gene encoded sequence
CC	(AAR80321) linked to the human fibrin cross-linking substrate peptide
CC	93.2 sequence (AAR80316). The polymers can be used in biological
CC	systems where in situ formation of a biocompatible material with
CC	structural integrity is required e.g. as medical adhesives and sealants
CC	or for wound closure or tissue repair.
SO	Sequence 762 AA;
Query Match	14.5%; Score 175; DB 16; Length 762;
Best Local Similarity	35.0%; Pred. No. 0.00014;
Matches 63; Conservative	8; Mismatches 51; Indels 58; Gaps 12.
Oy	58 ASGP-----MGDP-----PGPISLPPESTPCSP-LFGQOPADAAEPGKPTG 100
Dy	29 ASDPMGAPGLPSPGQGLPSPGAPGLTPGQPLP--GSPAAGTLPGQLPSPGAPLPG 86
Oy	101 SEDHQRATQLTGALDTSPTGHFFAFETP----IPGASMSISPTTPG--GLPW--GM 152
Dy	87 GQGQHILGARG--AGDVGPSG-----APGLTPGQGLPSPG--APGLTPGQGLPSPGA 136
Oy	153 TQTPTP-----VEDMDGGQGEDLRH-----PGISLPADLPGRPERCPHP 191
Dy	137 PTCTPPGQGLPSPGAPGLPGEQGQHILGARGAGDVGSAGAPGLTPGQGLPSPGAPT 196
RESULT	9
ID	AAR80324 standard; Protein: 762 AA.
XX	
AC	AAR80324;
XX	
DT	18-APR-1996 (first entry)



(PROT-) PROTEIN POLYMER TECHNOLOGIES INC.

Cappello J, Ferrari FA;

WPT; 1998-387004/33.

Recombinant collagen-like polymers - useful for making gels, films fibres, etc.

This sentence represents a unnatural collagen-like polymer of the invention. The products may be used as films, fibres, moulded objects and admixed with other natural or synthetic polymers or coatings on fibres, films, labware or other surfaces, e.g. prosthetic devices. The polymers may be used for binding a wide variety of specific binding materials, as catalytic substances (where the amino acid sequence may specifically chelate a wide variety of elements), as purification media, composites, laminates or adhesives. They may also be combined with inorganic or organic materials such as carbon fibres, nylon fibres, microcellulose, etc., as flask coatings or in synthetic matrices for the growth and study of cells, as affinity columns or as supports for biological materials. The polymers have collagen-like properties, but may

biological materials. The polymers have collagen-like properties, but may be easily expressed in micro-organisms in high efficiency. The new sequences can be tailored to give the desired properties.

Query Match	14.5%;	Score 175;	DB 19;	Length 762;
Best Local Similarity	35.0%;	Pred. No. 0.00014;		
Matches 63; Conservative	8;	Mismatches 51;	Indels 58;	Gaps 12

[illegible]

RESULT	13	.
AAW57668		
ID	AAW57668	standard; peptide; 762 AA

DT 27-AUG-1998 (first entry)  
 YY

Collagen-like polymer.

KM Collagen-like polymer; synthetic polymer; fibre coating  
KM prosthetic device; catalytic substance.

PN US5773249-A



OS	Synthetic.	
XX		
PN	US5773249-A.	
XX		
PD	30-JUN-1998.	
XX		
PE	02-MAY-1996;	9605-0642255.
XX		
PR	02-MAY-1996;	9605-0642255.
PR	04-NOV-1986;	8605-0927258.
PR	29-OCT-1987;	8705-0114618.
PR	09-NOV-1988;	8805-0269429.
PR	06-NOV-1990;	9005-0609716.
PR	12-NOV-1991;	9105-0791960.
PR	05-NOV-1992;	9205-0972032.
PR	22-DEC-1995;	9505-0577046.
XX		
PA	(PROT-) PROTEIN POLYMER TECHNOLOGIES INC	
XX		
XX	Cappello J, Ferrari PA;	
	WPI; 1998-387004/33.	

Recombinant collagen-like polymers - useful for making gels, films, fibres, etc.

PS Example 7; Column 51; 93pp; English.

CC This sequence represents a unnatural collagen-like polymer  
CC of the invention. The products may be used as film, fibres, moulded  
CC objects and admixed with other natural or synthetic polymers or coatings  
CC on fibres, films, labware or other surfaces, e.g. prosthetic devices. The  
CC polymers may be used for binding a wide variety of specific binding  
CC materials, as catalytic substances (where the amino acid sequence may  
CC specifically chelate a wide variety of elements), as purification media,  
CC composites, laminates or adhesives. They may also be combined with  
CC inorganic or organic materials such as carbon fibres, nylon fibres,  
CC nitrocellulose, etc., as flask coatings or in synthetic matrices for the  
CC growth and study of cells, as affinity columns or as supports for  
CC biological materials. The polymers have collagen-like properties, but may  
CC be easily expressed in micro-organisms in high efficiency. The new  
CC sequences can be tailored to give the desired properties.

**SQ Sequence 762 AA;**

Query Match	14.5%	Score 175;	DB 19;	Length 762;
Best Local Similarity	35.0%;	Pred: No	0.00014;	
Matches 63; Conservative		8; Mismatches 51;	Indels 58;	Gaps 12;

```

58 ASKP-----NPGED-----PGTSLPPESTGSP-LPGKODPADAEPPGPTG 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29 asqpmgaprlppbpqqlpsapapqrlcpbpqlp--gsppaprlppbpqqlpsapapqrlppbp 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 SEDHGGRATQLPGALDITSSPGTQRPAPPEP-----IPGASMSISPTTPG--GLPW--GW 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
87 gqghhlggarq--agdwspg-----apqlppbpqlpbspq--apqlppbpqlpbspq 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 153 TORPPT-----VPEKGQGGEDLRH-----PGTSLPADLRGAPPEPCHP 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 137 pglppbpqqlpsapapqrlcpbpqqlp--gsppaprlppbpqqlpsapapqrlpbspqapqlp 166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db

```

RESULT 14

ID AAW53535 standard; Protein; 762 AA.

AC AAW53535;

DT 10-AUG-1998 (first entry)

DE Amino acid sequence of the plasmid encoding fibrin gamma PPA1-A.

XX

KW fibrin gamma PPA1-A; peptide repeat unit; DNA repeat unit;  
KW high molecular weight polymer; synthetic silk; silk worm.

OS Synthetic.

PN WO9810063-A1

PD 12-MAR-1998.

PF 23-SEP-1996; 96WO-US15306.

PR 03-SEP-1996; 96US-0707237.

PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.

PI Cappello J, Crissman JW, Dorman MA, Ferrati RA,

DR WPI; 1998-193613/17.

PT Preparation of synthetic repetitive DNA - useful for construction of

PT material, e.g. synthetic silk

PS Example 6; Pages 83; 127pp; English.

This is the amino acid sequence of the fibrin gamma chain encoded by the  $\gamma A$  gene, which is used in the method of invention.

the preparation of synthetic DNA sequence having repeating about 3-15 codons and encoding a protein of at least about 30 kDa.

polymers (e.g., synthetic silk), either nucleic acids or peptides that

are the expression products of the nucleolar rDNA. These are the molecular weight peptides containing repeating units which are useful

XX  
XX

SQ Sequence 762 AA;

Query Match	14.5%;	Score 175;	DB 19;	Length 762;
Best Local Similarity	35.0%;	Pred. No. 0.00014;		
Matches 63;	Conservative	8;	Mismatches 51;	Indels 58;
				Gaps 12;

```

QY      58 ASGP-----NRPDP-----PGTSLSPSTSPSP-LPQPQPADAAEPKPGPTG 100
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      29 aadpmgagpqlpbgpbgqlpbgpbgatpbgqtlp--gspapqtpbgqlpbgspagpqtpg 86
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      101 SEDHGGRATOLPGLADITSSPCTQHPFAPEP-----IPGASWISIPPTPG--GLPW--GW 152
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      87 gqghhlgakq---agdvagspg---apqlpbgqlpbgspg---apqtpbgqlpbgspga 136
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      153 TQTPTP-----VPEDKGGCGEDLRH-----PGTSLPADLPGRPEFCIP 191
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      137 pqtppgqqlpbgspagpqlpbgqgqhhlbgakagaaqdvagspagpqtppbgqlpbgspgagpqlp 186
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15

ID AAG92860 standard; Protein; 635 AA

AC AAG92860

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 6614

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide

XX

XX

XX  
XX  
  
XXXX  
  
XXXXX

PD 20-JUN-2001

XX 18-DEC-2000; 2000EP-0127668.  
 XX  
 PF 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOSYO KK.  
 XX  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A,  
 XX  
 DR WPI; 2001-376931/40.  
 DR N-PSDB; AAH68079.

Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene

Claim 17; SEQ ID NO: 6614; 246pp + Sequence Listing; English

The present invention provides a number of nucleotide and protein sequences from the *Corynebacterium* *Corynebacterium* glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of *corynebacterium* *corynebacterium*, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from *Corynebacterium* *corynebacterium*, and identifying a homologue of a gene derived from *corynebacterium* *corynebacterium*. *Corynebacterium* bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention.

Note: The sequence data for this invention is available from the European Patent Office under the accession number EP 1 111 111.

sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

50 Sequence 635 AA,

Query Match	13.2%;	Score 159.5;	DB 22;	Length 635;
Best Local Similarity	27.8%;	Pred. No. 0.0015;		
Matches 71;	Conservative 13;	Mismatches 78;	Indels 93;	Gaps 15

```

03  / POSGPHLPINKISSHEPTGDR-----SYVCEMEVDRCSITG 45
04  56  pmduq -tspnktisttpppagnalpapqgalptprakteqdaivppptaaklpv-----pg 107
05  46  YHRLCVCSSICKKASPNG---PDP---GPTSLPPSPSPG---LEGPDPADAAP 94
06  108  -----sstipagralptcpvaggsvdaprasapavvnpaagaavpbaqsi 155

```

Db 156 paapaapsaalpptaalpvaasatpvpagvapaasapsalpvg--svrppagpisa 214

149 ppgwtotp-----fpvppdkqpced-----lrhpgslpa- 179

Db 213 pgsalptpgsaaplpvgalpvtggaalpvaasapsalpvgjlpvtgvpptgpaalpvp 272

QY	180	---	DLPGRPPEPCHP	191
Db	273	gapdapgtpsi	paap	287

RESULT 16

ID	standard; Protein; 682 AA.
AA80330	

AC AAR80330 ;

DT 18-APR-1996 (first entry)

DE Protein polymeric adhesion substrate 1-C.

XX	Pendent group; repeating unit; enzyme recognition site; sealant; fibrin
KM	enzymatic cross-linking; biocompatible material; structural integrity;
KM	medical adhesive; wound closure; tissue repair; transglutaminase.
XX	Synthetic.
OS	
PM	
PM	
XX	W09523611-A1.

PD 08-SEP-1995  
XY

PF 03-MAR-1995; 95WO-US02728.  
VY

PR 03-MAR-1994; 94US-0205518.  
XY

PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.  
XY

PI Cappello J;  
YY

DR WPI; 1995-320413/41.  
XX

PT Protein polymers comprising repeating units and sequences - capable of enzyme-catalyzed coupling to a bound format

PT biocompatible material for wound closure and tissue repair  
XX

Example 4; Page 47; 138pp; English  
xx

The amino acid sequence of the protein polymeric adhesion substrate (PPAS) 1-C, the protein comprises 10 repeats of the PPS1-C monomeric repeat (AA8080329) which consists of the CIP 3-7 gene encoded sequence (AA8080321) linked to the human fibrin cross-linking substrate peptide 33-4 sequence (AA8080317). The polymers can be used in biological systems where in situ formation of a biocompatible material with structural integrity is required e.g. as medical adhesives and sealants or for wound closure or tissue repair.

SQ Sequence 682 AA;

Query Match	13.1%;	Score 158;	DB 16;	Length 682;
Best Local Similarity	37.6%;	Pred. No. 0.002;		
Matches 62; Conservative	9;	Mismatches 50;	Indels 44;	Gaps 14

```

QY 58 ASGP-----NPGPD-----PGPTSLPPESTPGSF-LPGQDDPADAAEPPGKPTG 100
    |||      |||      |||      |||      |||      |||      |||      |||
Db 29 ASDPMGAPGtLpGPqgIpgSPpGApGtLpGpGqIh--gSPGApGtLpGpGqIpgSPpGApGtLpG 86

```

QY 101 SEDHQHGRATQLCALDITSSPGTQ---HPFAETP----IPGASWSISPTTPG--GIPLW 150  
:: | | | | | | | | : : : | | |  
Db 87 ak-qagdvgs-pgapgt--pppqg]lpspspgapqtpqpqglpspg--apqtpppqg]lpq 139

QY 151 -GWTGPTIPVPE--DKGQPEDLRHFGTSLPADLPGRPEPCAP 191  
 - - - - - : - - - - - - - - - - - - - - - - - -  
 Db 140 spgagptpggagkagdgvspra----ptprpqglpgspapgtpr 180

RESULT 17

AAW49717 standard; Protein; 682 AA.

AC AAW49717;

DT	12-OCT-1998	(first entry)
XX		

Protein polymer adhesive substrate PPA1-C.

Protein polymer adhesive substrate; PPA1-C; sealant; wound healing; CLP 3.7; collagen; human.

Synthetic.

US57773577-A. PN



OS	Synthetic.	
XX		
PN	US5773249-A.	
XX		
XX		
PD	30-JUN-1998.	
XX		
XX		
PF		
XX	02-MAY-1996;	96US-0642255
PR	02-MAY-1996;	96US-0642255
PR	04-NOV-1986;	86US-0927258
PR	08-OCT-1987;	87US-0114618

PR 06-NOV-1990; 90US-0609716.  
PR 12-NOV-1991; 91US-0791960.  
PR 05-NOV-1992; 92US-0972032.  
PR 22-DEC-1995; 95US-0577046.  
XX

XX (PROT-E) PROTEIN POLYMER TECHNOLOGIES INC.  
XX  
PI Cappelletto J., Ferrari FA;  
XX  
DOI: 10.1002-2872004/23  
DE

Recombinant collagen-like polymers - useful for making gels

XX  
XX  
PS  
XX

lures, etc.

Example 5; Column 45; 93pp; English.

This sequence represents a unnatural collagen-like polymer of the invention. The products may be used as films, fibres, moulded objects and admixed with other natural or synthetic polymers or coatings on fibres, films, lapware or other surfaces, e.g. prosthetic devices. The polymers may be used for binding a wide variety of specific binding materials, as catalytic substances (where the amino acid sequence may specifically chelate a wide variety of elements), as purification media, composites, laminates or adhesives. They may also be combined with inorganic or organic materials such as carbon fibres, nylon fibres, nitrocellulose, etc., as flask coatings or in synthetic matrices for the growth and study of cells, as affinity columns or as supports for

biological material

XX  
SO Sequence 417 AA.

sequence 4.1 / AA;

Query Match	12.8%;	Score 154;	DB 19;	Length 417;
Best Local Similarity	35.9%;	Pred. NO. 0.002%;		
Matches	60;	Conservative	7;	Mismatches 46; Indels 54; Gaps

29 asdpmgagptgpgqglpspsagptgpgqlp--gspsagptgpgqglpspsagptgpg 86  
 Db

```
QY ---TGSSEHQHGRTATQLPGALDITSSPCG---HPFAPEPP---IPCASWSISPTTPC- 14
      ||| | | | | | | | | : | | |
Db    87 qgIpgS-----pgapgt--pqpqIpgSpgapagtbpqpgIpgspg--apgtppp 13
```

Db 132 991p9sp9ap9t9p9q91p9sp9a-----p9t9p9q91p9sp9ap9t9p 174

Search completed:

Job time: 5535 sec

Thu Aug 29 10:46:17 2002

us-09-938-330-22\_copy\_1040\_1249.rag

Page 12

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 20:06:58 ; Search time 53.85 Seconds  
(without alignments) 2234.055 Million cell updates/sec

Title: US-09-938-330-20

Perfect score: 6954

Sequence: 1 MAPRLALSYLLPLHCAALCT.....PTGPTLGLVLRPSQLRGHT 1252

Indexing table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 71.1\*\*  
2: PIR1\*\*  
3: PIR2\*\*  
4: PIR3\*\*  
5: PIR4\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3326	47.8	1205	2	procollagen N-endo
2	1108.5	15.9	951	2	gene ADAMTS-1 prot
3	1072.5	15.4	2165	2	hypothetical prote
4	896.5	12.9	837	2	hypothetical prote
5	774.5	11.1	550	2	hypothetical prote
6	699.5	10.1	1558	2	hypothetical prote
7	699.5	10.1	1558	2	protein C37C3.6a I
8	677	9.7	1444	2	hypothetical prote
9	622	8.9	951	2	angiogenesis inhib
10	556.5	8.0	860	2	hypothetical prote
11	472.5	6.8	957	2	hypothetical prote
12	433	6.2	898	2	hypothetical prote
13	430	6.2	1059	2	hypothetical prote
14	420	6.0	826	2	hypothetical prote
15	398.5	5.7	571	2	monocyte surface a
16	361	5.2	609	2	catrocollastatin P
17	339.5	4.9	616	2	ecarin precursor
18	312	4.5	549	2	metalloproteinase
19	306.5	4.4	776	2	androgen-regulated
20	302	4.3	610	2	vascular apoptosis
21	294.5	4.2	617	2	metalloproteinase
22	294.5	4.2	814	2	disintegrin-like m
23	294	4.2	1074	2	semaphorin F precu
24	276.5	4.0	1584	2	brain-specific ang
25	273.5	3.9	478	2	fibriolytic metal
26	272	3.9	903	2	metlin alpha - mo
27	269	3.9	903	2	fibrolyase (EC 3.4.
28	266.5	3.8	1042	1	hypothetical prote
29	266	3.8	789	2	androgen-regulated

Noting  
C105  
E10096

#### ALIGNMENTS

30	264	3.8	777	2	I48100
31	257.5	3.7	480	1	A30065
32	255	3.7	823	2	S18968
33	252.5	3.6	416	2	A37877
34	248	3.6	481	2	JC4342
35	245	3.5	429	2	A42972
36	240.5	3.5	905	2	S55059
37	240	3.5	407	2	S62600
38	240	3.5	478	2	JQ1301
39	238	3.4	1170	2	A40558
40	237.5	3.4	481	2	S43125
41	235.5	3.4	1170	2	TSHP1
42	234	3.4	660	2	S71949
43	234	3.4	734	2	JC4861
44	233.5	3.4	419	2	S41607
45	232	3.3	735	2	G02937

RESULT 1  
118517  
procollagen N-endopeptidase (EC 3.4.24.14) I - bovine  
N:Alternate names: procollagen N-proteinase  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18517  
R:Collage, A.; Nusgens, B.V.; Lapiere, C.M.  
submitted to the EMBL Data Library, February 1996  
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.  
A:Reference number: Z18941  
A:Accession: T18517  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1205 <COL>  
A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CA65253.1  
A:Experimental source: skin  
A:Genetics:  
A:Gene: PC I-NP  
A:Function:  
C:Description: catalyzes cleavage of the propeptides of type I and II collagens prior  
C:Keywords: hydrolase; metalloproteinase

Query Match 47.8%; Score 3326; DB 2; Length 1205;  
Best Local Similarity 52.1%; Pred. No. 1e-188;  
Matches 641; Conservative 160; Mismatches 351; Indels 78; Gaps 20;

QY	7	LLSYLLPLHCAALCTAGSRTPELHLSGKLSIDGVYFGCTDPRGRFLSHVYSGPAASAG	66
DB	17	LLLLPLPADARLAAADDPGGHGAERTLAVPVTDAGLVSHVSA-ATAPAG	75
QY	67	SNVVDPTPLPRHSHSLRVARSPLHPCGTLMPGGRHSLSYFNVPFCKELRLRPPR	126
DB	76	---VTRRRAPAIPIPLS-GGSEEDPG-----RLFNIVVFRRDLHLRLPPAR	121
QY	127	LVPVGSSEVMEQDFELFROPRLRQCVYTGVTGM-PAANAISNCDLACILRTDSTDF	185
DB	122	LVAAGTAVEMOGESCATRVEPLGTCLYGDVAGLAESSVALSNCDLACILIMEEERF	181
QY	186	FIEPLRGGQOEKA-SGRHVYVYRRRAVOQEAERDGS-----DLHNEAFGGLDPLNLGL	239
DB	182	FIEPLRGGQOEKA-SGRHVYVYRRRAVOQEAERDGS-----DLHNEAFGGLDPLNLGL	239
QY	240	VGDGLDPTERK-RRHAKPSYSIEVLAVVDSVYAFHKEHVQNVYLTIMNIVDEIYHDE	298
DB	240	LEERVNYSRRRRRRRAADDVYIEVLGVDSVYQFHTENHVKTLITIMNIVDEIYHDE	299
QY	299	SIGVHNTNLVALLVNGVROSLSIERGNPERSLSBOVCRAHNSOORDPSHAHHHVF	358
DB	300	SIGAHINVLVRLILLSTGKSKSLIEIGNPSOSLENVCRAVYLOOKBDPTDHDVHDHAF	359





Db 814 -----HALPKIKFTYF-----MKKTESF-- 833  
 Oy 900 RCNHPGSPYVWTEWEGSCSGKIGVOTRGIOCLPLPSNGHKWPKAKACGD-RPE 958  
 Db 834 --NAIPTSE-WVDEWEGSCSGS-GWOKRVQC--RDING--HPSACEAKEVPA 883  
 Oy 959 ARRCPLRVCPAOWRLGAMSOCATGEGIQOQOVYCRTNANSLGHCEGDRP-----DTY 1013  
 Db 884 SFRPCADLPCE-HWQVGMSPSCSKTCCKGKYKRLKCVSHDGVLSNECDPLKKPKHY 942  
 Oy 1014 QVCSLPAC 1021  
 Db 943 DECTLTOC 950

UNT 3  
 71  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T21371; T24896  
 R:Gaiadety, S.  
 submitted to the EMBL Data Library, February 1996  
 A:Reference number: Z19413  
 A:Accession: T21371  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2165 <WIL>  
 A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3  
 A:Experimental source: clone F25H8  
 R:Gaiadety, S.  
 submitted to the EMBL Data Library, February 1996  
 A:Reference number: Z19949  
 A:Accession: T24896  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2165 <WIL>  
 A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3  
 A:Experimental source: clone T13H10  
 C:Genetics:  
 A:Gene: CESP:F25H8.3  
 A:Map position: 4  
 A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 87

Query Match 15.4%; Score 1072.5; DB 2; Length 2165;  
 est Local Similarity 26.0%; Pred. No. 3.2e-55;  
 Matches 312; Conservative 176; Mismatches 425; Indels 287; Gaps 45;

Db 151 ECVYTGCVTGMPCGAANAISNC--DGLAGLIRTDSTDFIEPLERG--QOEKESGRTH 204  
 Db 168 DCYIRAHVKVGHOSH-VNLCSEDELGYGLMPLSGIHTEVERPLISNGTEHDCARRHQH 226  
 Oy 205 VVYRREAVOGEAEPDGLAHNEAFGIGDRLNLLGLVGQDLGT-ERK--RRNAKPGSYS 260  
 Db 227 LVKRPDMNRKSD-----HLNSTVNETETTVATWQDMEDVIERKARSRAAASMDHY 281  
 Oy 261 IEVLAVVDSVVFHCKEHVQNYVLINLVDEIYHDESLGVHINIALVRLIMWGYRSL 320  
 Db 282 VEVLVAVADTKMEYHGRS-LEDYVLILFSTVASITRHOSLRASIVVYKLLIVLKTENAG 340  
 Oy 321 SLIERKNPERSLEOVCRMAHSSQORODPSIAEHHDHIVFLTRQDFGSPS-----GYAFVT 373  
 Db 341 PRITLO-NAOOLTDPCFRWQOYVNDPSSSVQHNDVAILLTRKIDLCRSOGKCDTLGLAELG 399  
 Oy 374 GCHPLRSCALNHEDFSFAFVIAHETGVHLEMEHDGOGNCA-----D 417  
 Db 400 TWCMDQKSCALIEDNLSAFLIAHGLGVFSIPHDDE-KKCSITMPVKNKYCKQSTKFD 458  
 Oy 418 EYSLGS---VMAFLVOAARHHRMSCKSLRYLPS---YDCLLDDPDPAMPQP-- 468  
 Db 459 KTOFONNHIMAPLTLEYNTHPWSWSPSCSAGMLERFLNNRQOTCLFDQPYERRYEDVF 518

Oy 469 -PELPGITVSMDEQCRDFEGSGVOTCLAFRTFEPCKOLWC-SHPDNPFCYCKTKKGPBLDG 526  
 Db 519 VADEPKRYADAHQCKEFVGPASIELCPYMP--CRRLMCAFEYSGOMGCRTOHMPADG 575  
 Oy 527 TECARGK--WQFKGCIKMSPEQTYGDDGSSWTKRSGSRGCGGVYRBSRSCNNPSP 584  
 Db 576 TPDCBSRSMFCHHAGVCLAPESLTKIDGQGMWRSMGECSPKCGGVLGDLCDSPAP 635  
 Oy 585 AYGRPCGLPMFEYOVENSECPCTYEDFPAOQCAKANSYVH-----QNAKHSWP-YEP 639  
 Db 636 RRGKCYGQREYRSCNTQCECPMDTPYREVYCSERNNDIGLQVASTNTHTWYKYXAN 695  
 Oy 640 DDDAKCELTQSDADTGVVFMNOVHDTGRCSTRDYSYCARGECPVPGCDKEVGSMA 699  
 Db 696 VAPNERCKLYCRISGSAFYLIRDKYVDGTPCD-RNGDDICVAGACMPACGCDHQLSTLR 754  
 Oy 700 DDKGCVGGDNSHCRTVYKGLTGKASKAGALAKIYQIPAGARHIOIAL-----EKSHRSV 755  
 Db 755 RDKGCVGGDSSCKYVKGTFNEDG-TFGYNEWKIIPAGSANIDIKGYNNKKEDDNTL 813  
 Oy 756 VKNOVTSFLLNPKGKEATSR--TFPMGLEW-EDAVEDAKESLKTSGPLPEALAILA 810  
 Db 814 SLRANGFELLNGHFQVSLARQOIAFDVLEYSQDAI--IERINGTPIRSDIYHV 870  
 Oy 811 LPTEGGRSSLATKY-----YIHEDLLPLIG-----SNNVLE 844  
 Db 871 L-SVGSHPPDISEYMTAIVPNAVIRPISALYLRVITDVTWTECDRACGQOOSOKIMCL 928  
 Oy 845 EMDY-----EMALKSNAPSCAKGGIGTFTXGCR 877  
 Db 929 DMSHHRSHDRNCONVLRPKQATRMKNDICSTWITTEDVSSCAKCSGOKHQRVSCVM 988  
 Oy 878 RDHNM--VQNHLDCHKRPPI----- 897  
 Db 989 EGDROTPESEHLDRNSKSPDIASCYIDCSGRKMYNGWETSCSETGSGNGKHKRSYCDV 1048  
 Oy 988 -----RRCONHPGSPYVWTEWEGSCSGKIGVOTRGIOCLLP 938  
 Db 1049 DSNRRVDESICGEGQKATEBECNRLPC--PRVYGHMSECSRSDG-GVKRHAQOCIDA 1105  
 Oy 939 LSNQTH--KYMPAK-----AC-----RPE 958  
 Db 1106 ADRETHSRGPAQOTQEHCHENACTWMOFGVMSDCSAKAGDVOYRDANCIDRHSVLP 1165  
 Oy 959 AR-----RCLRVCPAOWRLGAMSOCATGEGIQOQOVYCRTNAN-----SLGH 1004  
 Db 1166 HRLCKMEKTIITKPCRHRESCP-KYKIGEMSQSVSGEDCSRRVSCVSGNGTEYVMSLGC 1224  
 Oy 1005 CEGDRPDTVOVSL-----PACGNHONSTVADVWELGTETEGOVVPSG 1049  
 Db 1225 TASDRASHOTCMLGTCPPWRNTDMSACVSGIGHRETTTC-TYRQSVASFCGDT- 1282  
 Oy 1050 PLHPINKISSMCAEPCTGDR-----SVFCOMEVLDRYCSIPGHRILCVSCIKKASG 1102  
 Db 1283 ---KMPETSQTHLPLCTSMKPSHMSPCSVTGSGLQGR-----SVSCTRSGEG 1328

RESULT 4  
 T00355  
 hypothetical protein KIAA0688 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
 C:Accession: T00355  
 R:ishikawa, K.; Nagaoka, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, D.  
 DNA Res. 5, 169-176, 1998  
 A:Title: Prediction of the coding sequences of unidentified human genes. X. The compl  
 A:Reference number: Z14142; MUID:98403880  
 A:Accession: T00355  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-837 <ISH>  
 A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BAA31663.1; PID:g3327190  
 A:Experimental source: brain



```

RESULT      7
T34395
hypothetical protein C37C3.6b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000
C:Accession: T34395; T34394
R:Geiselt, C.; Bradshaw, H.
submitted to the EMBL Data Library, July 1996
A:Description: The sequence of C. elegans cosmid C37C3.
A:Reference number: Z21518
A:Accession: T34395
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2167 <GET>
A:Cross-references: EMBL:U64857; PIDN:AAC25866.1; GSPDB:GN00023; CESP:C37C3.6b
A:Experimental source: strain Bristol N2; clone C37C3
A:Accession: T34394

```

```

QY 551 OGGMSWTKFEGSCRCSCGGVRSRSCNNPSPAYGCRPLGPMFEYUQNSNEEC -PGT 609
Db 75 EFGNMGPWVPEMECSCRCGGGQVLEKRCSCG-----DCTGASVATYISCLNACESGT 126
QY 610 YEDFRAOQCAKRNSTYYVHQNAKHSWVPYEPDDAOQKCELICQSDADTGDVVF--MNOVVD 667
Db 127 --DFRAEQCSKFPNDALDGN-YHKMTPYK--GKNKEIWCQ-PESONFYYKADVV-D 178
QY 668 GTRCSTRDPYSCANGCEVPPVGCCKEYSGSMKADRCGVCGSDNSHCFTVGTLGAKSKQA 727
Db 179 GTRCQSKS-NDICVNDGECILPVCGBCKLSSLKTFPKCKCDGSDTCTTEGFRDERNLSP 237
QY 728 GALKLVQIPAGARHIQIOLALEKSPHRYSVQYQVGS--FLTNPKGKATYSRTTLAG--L 783
Db 238 GHDIKILKEPAGATNIKKIDAKRSNNIUALKN--GSHFPLNGLNGILOVEKEVEVGITF 294
QY 784 EWEADVEDAKESLKTGSPLEPAIALALPREGGPRSLAKYVIEHDDLPLTGSNNVL 843
Db 295 VYDDA---EPETLSAQGBLDELTV-ALLTRKSGROIAIYEE-----STEL 337
QY 844 EEMDYYEALISMAPCSKACGCGGIQFTKTCGRRRRDHMYORHLCHKKRPK-IRRCN 902
Db 338 EEEVDYMKFEQNMTPCYSVSCGKGYQTNLYCIBGKNNGRVEDDLCEENATKPEFEKSCE 397
QY 903 QHPCSQPPWVTEEMGQASCRCSCGKIGVQTRGICOLLPLNSNTHKVMYAKKACADDEARAP 962
Db 398 TYDC-EALMTLGTDMESGSSSTCGDGOQOYRYVYVYHQVFANRRVTVEDGNCVTERPPYQOT 456
QY 963 CLRVPCPAQWMLGAMSOCASATCGEGIDQOROVGRTN-----ANSIIGHCEEDRPT 1012
Db 457 CNRFACP-EMQAGPMSACEKCGAPFYRSVTRORSKEKEBEGKLLAADAACPADDEBERDPT 515
QY 1013 VQVQSLPACGSGNHONSTVRADWEL-----GTPBGQWVPQSGPLADINKI----- 1057
Db 516 ERTCNLGPCEG---LTFVTJGEMNLCTRNDTEETREVLCKDSOGGRAYPLEKCLVDNSTE 571
QY 1058 -----SSMKAAP-----CTGD-----RSVFCQ-----MEYDR-XC 1083
Db 572 IPTOTRCAQTQPCREYEMTWSEMSKCTTEGCGHGTKTRRYVALAHONGEILEVDESHC 628

```

RESULT 8  
T18856  
angiogenesis inhibitor homolog - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #extl\_change 18-Feb-2000  
C:Accession: T18856; T24653  
R:McMurray, A.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: Z19031  
A:Accession: T18856  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-144 <MIL>  
A:Cross-references: EMBL:Z50004; PIDD:CAA90293.1; GSPDB:GN00028; CESP:C0ZB4.1  
A:Experimental source: clone C0ZB4  
R:McMurray, A.  
submitted to the EMBL Data Library, July 1995

A:Reference number: 219917  
 A:Accession: T24653  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1444 <M12>  
 A:Cross-references: EMBL:250006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1  
 A:Experimental source: clone T07C5  
 C:Genetics:  
 A:Gene: CESP:C02B4.1  
 A:Map position: X  
 A:Insertions: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3; 566

Query Match 9.7%; Score 677; DB 2; Length 1444;  
 Best Local Similarity 24.9%; Pred. No. 4.7e-32;  
 Matches 285; Conservative 135; Mismatches 429; Indels 296; Gaps 63;

QY 76 LPRHSHLTVARSPHPCGFLMPGRVGRHSILFNVTFGKEHLRLRPNRRLVPG--SS 133  
 26 VPEYSL-TEATRHPLKNG-----NLKKFTAMMDYHLNKRKRIVSPHIIISV 73  
 134 VEMOEDFRELFPOLR--QECVYTGVTGMPGAANAISNCGDLAGLITDSTDFIEPLE 191  
 74 VRHGDVVITY-AGLRDEQCHTQGEVKSNGNKAISDCAIMSGIYVEDHFLVLOTLP 132  
 QY 192 RG---QOEKASRTIVYVRRRAVOEAEFPGDLDHNAFGLDPLNLGLVGDOLD- 246  
 133 KRVHLLKER-----HLVYKRSAGLITNAE--SKTREETRLQEOESFCTDSEOLDP 184  
 QY 247 -----TERKRRHAKPGS-----YSIEVLVVDVSVFHCKEYHON----- 282  
 185 AMTIPALHFRVYITPTSAQDSSFTIPNMDPITLLEIGLFLDSLFHFREXYTODAEHL 244  
 QY 283 --VYLTMINIVDEYHDESLGVHINIALVRLVWGTROSLSL-----ERGNPSRLEBY 335  
 245 LEFSIALINNVHVLXOODTLPLNDIYVREYEM--WRTOPSALSTGVHNGQAQSLIDAF 302  
 QY 336 CRM-AHSQQRDPSEHAEHNDHVVYLRQDP-----GPSGAPTYGCHPLRSCALNHEHG 389  
 303 CRYQHMHNPSTDLDMNHIDYGVLLTYDLYHTTSVAGAPARAKCDPLFACSLVEGLH 362  
 QY 390 FSSAFVIAHETGAVLGEHDOGNGCADETSLSVANAPLVQAAFRHFHMSRCSKLELSRY 449  
 363 LGRSVFLAHENGMHMGVHDOVQOC---NKGCCILMSAVNGA--GRTYSDSCVREFNAF 417  
 QY 450 LPSTD-----CLDDPDPDPAW-----QPEELPGIYMSDEOCRRDFSGYGTCL- 494  
 418 LLOIDESGRGNCLRD---ASPLISTNHLSDILRLPGRFADQOCSTFGWDYVEIR 472  
 QY 495 -AFRTFECQOLWCSHPDNPFCCTKKKAPPLDGTGECAPGWCFFKGHCI--WKSPEQTYGO- 551  
 473 NGKAMDICRLIMCNSSTI---STAHPALESGWCANMKWCKGQCTH-----TFGLT 524  
 QY 552 ---DGNSSWTKFGSGSRSC-----GGG--VRSRSRSCNPNPSPAYGGRPLGPFMEY 598  
 525 PVPIDGEMSEW---GGAKEGCPIDQCAVSGSITVQGOHRCVNPAPNMGCKTEGANING 581  
 QY 599 QVC--NSEPCG-TYEDPRAQCAKRSYVYHONAKHSHWVPEPDD-----DAOKC 646  
 562 IYCGATSSNCLGFTREEGNKICS-----SIKTD--PHKPDQOLTEGFEHSTQRC 630  
 QY 647 ELICQADTGVVFMNQ--VVDGTGRCYSRDPYSVARGECVPVGCDEKSGVSMKADKCGV 705  
 631 RWCMLH--IGSFLRNKQFPDGTGCP-DAV--CVGGGCLALSCDKALVEQPED--CPR 684  
 QY 706 CGGDNH-----CRIVKGLGKASKOAGALKVLPAGAKRHQIDALEKSPHRSY 755  
 685 IEGRSVHOMEMSWSECSVSCGLGGERVERKRC-----SSGRCKGVSESRPEGV 737  
 QY 756 VKNQVTSFLNPKKEATSRPTAMGLEWEDAVEADAKESLKTSGPLPEAIIATLAPTE 815  
 738 L-----RDCEBFG-EW-----KEMGSCSEKCAL----- 759

QY 816 GPRSSIAKYVYIHEDLLPLIG---SNNVLLBEMDT-----YEWALKSNAPCSKACGG 866  
 760 -----GVOKRFRRCCLTDQSSKHLQEEPCDNEGCMTW--DEWSSCSQSCGG 806  
 QY 867 IOFTKICRRRDDHMHVQRLCDHKRRKPIRRCONHPCSQPVMYEWACGRSGKTL 926  
 807 ---RRYRIRKCLDDK---CDGDLEK---ESQNTQKCTISQSH--GDWLPCSVSG-I 851  
 QY 927 GVOTRGICQLPLSLNCTHVMKPAKACAGDRPEARRPCLRVPCQWRLGAMSQ-----C 980  
 852 GFQIR-----EKLCEGLCATANKKA-RTCNOQCCPSAFSLSVSEMGEMWTTC 898  
 QY 981 SATCGEIOOROVYCRTPNANSLHCEGDRPDTVQYCSLPACGNGHONSTVADWELGTP 1040  
 899 SATCGEGLQSRERSCR-----GCTEDDASQTRRCVNGPC--EHSYLT-----W----- 941  
 QY 1041 EGWVQSGPLHPIKNTSSMCAAEPTGDRSVCQMEYLDRCSPGYHRLCVCSCI--K 1098  
 942 -SEWT---TCETCSSFDSRRKRIACDG--TTENCODKIDEETCDI-----ACLREK 986  
 QY 1099 KASGP 1103  
 987 HSTGP 991

RESULT 9  
 T00260  
 hypothetical protein KIAA0605 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
 C:Accession: T00260  
 R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, DNA Res. 5, 31-39, 1998  
 A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp  
 A:Reference number: 214086; MUID:98290545  
 A:Accession: T00260  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-951 <NAG>  
 A:Cross-references: EMBL:AB011177; NID:93043733; PIDN:BA025531.1; PID:93043734  
 A:Experimental source: brain  
 C:Genetics:  
 A:Note: KIAA0605  
 C:Superfamily: thrombospondin type 1 repeat homology  
 F:46-106/Domain: thrombospondin type 1 repeat homology <THN1>

Query Match 8.9%; Score 622; DB 2; Length 951;  
 Best Local Similarity 24.6%; Pred. No. 5.2e-29;  
 Matches 204; Conservative 69; Mismatches 267; Indels 290; Gaps 29;

QY 552 DGG-----WSSWTKFGSGSRSCGGGVRSSRSRSC---NNPSPAYGGRPLGPFMEY 600  
 40 EGGTDATAFWMWGEWTKTAFRSRSCGGGVTQERHCLQORRKSVPGRNRTCTGTSKRYOL 99  
 QY 601 CNSECEGTVDYFAOOCARNSYVYHONAKHSHWVPEPDD-----DAOKCELICQADTG 656  
 100 CRVCEPDPGSRFEEQCVSFNS--HYNRTGHQMKPLIPDYVYHNSKFCDCILCTTVD-G 157  
 QY 657 DVVFMNQVHDGTGRCYSRDPYSVARGECVPVGCDEKSGVSMKADKCGVCGGDNHSHCRTV 716  
 158 QROLM-VPARGTGCKLTLDRGVCYSKCEPFGICGDDVLTSTHLDKCGICQDGGSSCTHV 216  
 QY 717 KGLGKSKQAGALKVLPAGAKRHQIDALEKSPHRSYVKNQVTSFLNPKKEATSR 776  
 217 TGNVRKGNANLGYSLVTHIPAGARDIIVRKKSAVDLALADE-AGYFFNGNRYKVDSPK 275  
 QY 777 TFTAMGLEWF-----DAVEDAKESLKTSGPLPEAII-----TALPPT 814  
 276 NFNIAGTVVYKRRPDVYETGIEYVAGPPLNGLNVMVNNQKSPSITFEYTLLOPH 335  
 QY 815 E----- 815

Db 336 ESRPPIYGFSESAESQGLDAGIMGFIPHNGLYQAGASSERLGLDNLRLFGHPGLMEL 395  
 QY 816 ----- 815  
 Db 396 GPSQGEITNEVCEQAGGACGCEPPRGKGRNRNVTGPTLGDKDEVDTHFASQEPFSA 455  
 QY 816 -----GGPRSSLAIVYI-----HEDLLPLTGSNNVL 842  
 Db 456 NAISDQLAGSGLDKDFTLNEVNSIFAGAGPSSLSAEFFVDYEEBEGAPYLINGSYL 515  
 QY 843 -----LEEDMTYEMALXSWAP 858  
 Db 516 ELSSRVANSSSEADPEPVNSTLTSAGNRTTKARKTPRKARKOGVSPADMYRMKLSHSP 575  
 QY 859 CSKACGGGIGFTYKGRRRRDHMHVORHLCDHKRKPRTARRCQNHCCSPVAVTEWGA 918  
 Db 576 CSATCTTGV-MSAVAMCRRYDGEVDSDYCDALTRPEVHEFCAGREC-OPKNETSSWSE 633  
 QY 919 CSRSRGKIGVTRGICLLPLLSNCTHKVMPAKACAGD--RPEARRECLRPVCPAOWRLG 975  
 Db 634 CSRTGGE-GYGFRRVRCMKMLSPGSDSSVSDLCFAEAARPEERKTCRNPACGPGWEMS 692  
 QY 976 AMSOCATGGE-GIQOROV-----VCRTNANSIG--HCEG---DRPDYOVCSLIPACG 1022  
 Db 693 EMSECTACGSRVVTDRDRCSEDEKLCDPNTRPVGEKNTGPPCDPQWTVSDWG-PCSG 751  
 QY 1023 GNRHNSYVRADYVELGTPEQOWYPOG-----GPLHPINKISSMCAEPCTGDRSVFC-- 1074  
 Db 752 SCGGGRIRHYCK--TSDGVRVPESSOCOMETKPL-----AIHCQ-GKKN--CRA 796  
 QY 1075 -----QMEVLDRCYSIPGHRCLCVSCIKKASG---PNEGPPDPGPTSLPP 1116  
 Db 797 HMLAQDERCNTTCGRGVKKRL-VICMELANCKPQTRSGPECGLAKKP 844

RESULT 10  
 T16892  
 hypothetical protein T19D2.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T16892  
 R:Bentley, D.  
 submitted to the EMBL Data Library, December 1995  
 A:Description: The sequence of C. elegans cosmid T19D2.  
 A:Reference number: Z18599  
 A:Accession: T16892  
 Status: preliminary; translated from GB/EMBL/DBJ  
 Molecule type: DNA  
 Sequences: 1-860 <BEN>  
 Cross-references: EMBL:U42846; NID:g1125809; P1D:g1125810; PIDN:AAA83600.1; CESP:T19D2  
 C:Genetics:  
 A:Introns: 56/3; 96/1; 171/3; 199/3; 245/2; 293/2; 335/3; 395/3; 426/3; 484/2; 505/3; 58

Query Match 8.0%; Score 556.5; DB 2; Length 860;  
 Best Local Similarity 24.9%; Pred. No. 3.4e-25;  
 Matches 227; Conservative 145; Mismatches 360; Indels 181; Gaps 52;

QY 168 ISMCDLGLIRTSTDFIEPLERGOQKEASGRTHVYRR--EAVQOEWA---EPDG 221  
 Db 55 VSN-----GMVYADNGDYHL-----IQSHQGTG-HVTHKRSIESIDHECOFANEDPY 102  
 QY 222 DLHNEAFGLDLPNLLGLVGDQGLGTERKRRHAKP---GSYSIEVLLVVDSDV---VRF 274  
 Db 103 PEDRETIANKSM-LIRMKD-----IRAEPRRDIITVELAVPADAMMDHFKM 152  
 QY 275 HGR---EHVQNTVLTLMNIVDEIYHDESLGVYINIALVR---LIMVGYRQSLIERGNP 328  
 Db 153 YGKAARENNHTFIMAVVNNIDVLYTORLLQPRINIKIVYELLKNITPLIMARKHSGNDV 212  
 QY 329 SSSLEQVCWMAHSQORODESHAHHHVVFLTRQDF-----GPSGYAPVGMCHPLRSC 382

Db 213 DRLLAFCQYONEINPPDADPRHMDALLFSGYDLHRNGVKTIVAGYAPYKMGCSGVRSQ 272  
 QY 383 ALNHEDGFSSAFVIAHETGVLHMEHDQNGCAGDETSLG--SVMAPLYOAFRFRHMSRC 441  
 Db 273 TINELDGFQSVFYVTHMGHSLGMYHGD-----NECDLRCTIMSPVSGG--KTHMSQC 325  
 QY 442 SKLELSRLPSTDCLLDDPEDPAMPPOP-----ELPGINYSMDQC 482  
 Db 326 SVENMATVGH-----LGDDF-----RPPNCLQDASANEQRMVAFKSEPPGQLFTLDEQC 376  
 QY 483 RFDGSGYQOTCLA-FRTPEP--CKOLMCSHPDNPEFCTTKGPPLDGTIECAPGWCKRGHC 540  
 Db 377 ELTFEGCKHKLKGGQTNQNICQMWGNGEBG--VIRTAH-PALLEGTYCGFGMICROGOC 433  
 QY 541 IWSPEQTYGODGQSSW-----TKFGCSR--SCGGVGR-SRS-RSCNNPSPAYGRCPC 591  
 Db 434 VGSQ--QLMRVTVGGMSWINDRPAFTGCGRSCQCEIRQIIMNSIQNCPNPSNNGAPC 492  
 QY 592 LGPMFEYQVCSSECCPG--TYEDFPAQCAKARNSTYVHQAOKHSWPYE-----PDDAQ 644  
 Db 493 QGDEARGMVCHRDVQNGDSIENATRVCSRLRD-----ENA-----IPNTILSGEJMFEQA 544  
 QY 645 KCCELICQADTGDVVFNNQVVDGTRCSYRDPYSVCARGCEVPVG--DKEVGSMAADQC 703  
 Db 545 MCKIMCLISGSTNIRTVSN-FPDGAPCG--PGQYCIKGRCLILGCTTLAYSEADCP 600  
 QY 704 GYCCGDS-----ECRTYKGTIGKASKOAGALKLVQIPAGARHIQIEALEKSPHRSVYKQ 759  
 Db 601 SVLDQTTTPPMPIHVSVDQFAGKTNPY-----KEHKKTFPLNEMGMSVMSBC 648  
 QY 760 VTGSFIINPKGKERTSTFTAMGLEWEDAVEAKESLKTSGPLPEALIALPPEGGPR 819  
 Db 649 VT--YDCHTQGVAVRARCIA-GV-----CAGLREBOPCTRPCTGSERPPLTTPQ 697  
 QY 820 SSLAKYVIHEDLLPLIG--SNVILEMDITYEALXSNAPCSKACGGGIGFTYKGRRRR 878  
 Db 698 QTFERNRPI-----APLPRQTNMILRKVD--HWG--PWAGCSVTGCTGQK-----RRR 742  
 QY 879 DHNVVQRHLCDHKRKPRIKRCNQHPQSPVAVTEEMKACSRGCKLGVOTRGICLLP 938  
 Db 743 ENCTGQ-----ECAGETGCPVMOGSCRENKNTIEW--SOWSDCSVNGGE-GVOFRKRCFAA 795  
 QY 939 LSNGTHTKVMPAKACAGDR--PEARPCCL-RVPCPAOWRLGAMSQCSATCGCIGQORQV 994  
 Db 796 FCRG--KDSVVRNCGYQRCSEITTRRPLVNRGSLMTGW--SSWSSCSTKCGIGQRTRRR 851  
 QY 995 CRTNANSLGHCQG 1007  
 Db 852 CYQ-----GSCIG 859

Search completed: August 28, 2002, 20:12:20  
 Job time: 322 sec

Thu Aug 29 10:46:11 2002

us-09-938-330-20.rpt

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 00:02:39 ; Search time 48.87 Seconds  
(without alignments)  
412,907 Million cell updates/sec

Title: US-09-938-330-22\_COPY\_1040\_1249  
Perfect score: 1206  
Sequence: 1 PEGQWVPSGPHLPINKISS.....PTGTFITCVLPDSQLRGHT 210  
Indexing table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	211.5	17.5	1205	2 T18517	procollagen N-endo
2	180.5	15.0	839	2 T04859	extensin homolog F
3	160	13.3	319	2 T32250	hypothetical prote
4	156.5	13.0	309	2 T29293	hypothetical prote
5	155.5	12.9	317	2 T19143	hypothetical prote
6	154	12.8	577	2 T09024	proline-rich prote
7	151.5	12.6	210	2 B44984	collagen - nematod
8	151.5	12.6	286	2 S34665	collagen, cuticula
9	151	12.5	296	2 A31219	collagen, I - Caeno
10	150.5	12.5	316	2 T21314	hypothetical prote
11	149.5	12.4	301	2 T21314	hypothetical prote
12	149	12.4	325	2 T32248	hypothetical prote
13	148	12.3	1664	2 T18262	S-layer protein -
14	146	12.1	290	2 T24590	hypothetical prote
15	145.5	12.1	299	2 T20605	hypothetical prote
16	145.5	12.1	316	2 T19291	hypothetical prote
17	145.5	12.1	330	2 T05717	probable extensin
18	145.5	12.1	1255	2 T31065	diaphanous protein
19	145.5	12.1	1340	2 A39808	proteoglycan core
20	145.5	12.1	2327	2 T42630	aggrecan - bovine
21	145	12.0	297	2 T27525	hypothetical prote
22	145	12.0	299	2 T29956	hypothetical prote
23	145	12.0	392	1 T19086	salivary proline-T
24	145	12.0	392	1 T19086	collagen alpha 1(I)
25	144.5	12.0	313	2 T22828	hypothetical prote
26	144.5	12.0	356	2 T22827	hypothetical prote
27	144.5	12.0	888	2 S28791	collagen alpha 1(X
28	142.5	11.8	360	2 T37285	collagen dpy-2 - C
29	142.5	11.8	464	2 S22697	extensin - Volvox

30	142	11.8	295	2 A44984	collagen - nematod
31	141.5	11.7	299	2 T24833	hypothetical prote
32	141.5	11.7	316	2 S08169	collagen col-12 pr
33	141.5	11.7	316	2 S08170	collagen col-13 pr
34	140.5	11.7	266	2 T22706	hypothetical prote
35	140.5	11.7	299	2 T22705	hypothetical prote
36	140.5	11.7	1690	1 CGH01B	collagen alpha 4(I
37	140	11.6	403	2 S52796	collagen alpha 2(I
38	140	11.6	677	2 S23296	collagen alpha 2(I
39	139.5	11.6	435	2 T15143	hypothetical prote
40	139.5	11.6	515	2 F70904	hypothetical prote
41	139	11.5	298	2 UC1448	collagen col-34 -
42	139	11.5	337	2 T23794	hypothetical prote
43	138.5	11.5	228	2 A44982	collagen uCOL1 - p
44	138.5	11.5	302	2 A31921	collagen dpy-13 pr
45	138.5	11.5	314	2 T32985	hypothetical prote

#### ALIGNMENTS

RESULT 1  
T18517  
procollagen N-endopeptidase (EC 3.4.24.14) I - bovine  
N:Alternate names: procollagen N-proteinase  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C/Accession: T18517  
R:Collge, A.; Nussgens, B.V.; Lapiere, C.M.  
Submitted to the EMBL Data Library, February 1996  
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.  
A:Reference number: Z18941  
A/Accession: T18517  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1205 <COL>  
A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1  
A:Experimental source: skin  
C:Genetics:  
A:Gene: PC I-NP  
C:Function:  
A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior  
C:Keywords: hydrolase; metalloproteinase

Query Match 17.5%; Score 211.5; DB 2; Length 1205;  
Best Local Similarity 29.1%; Pred. No. 2.2e-06;  
Matches 52; Conservative 24; Mismatches 68; Indels 35; Gaps 4;

QY 4 OMWVPSGPHLPINKISSSTEPICGDSVFCQMEVLDRTCSIFGHRCCVSC----- 54  
DB 1037 OMWVPSGPHLPINKISSSTEPICGDSVFCQMEVLDRTCSIFGHRCCVSC----- 54  
QY 55 IKKASGPNPGRPPPTSLPPESTPGSPPL---PGPODPADAEPGKPTGSDHGHGRATQ 111  
DB 1097 DDPRAEPPSGKNDIDELMPTLSVPTLVMEVQPPPGICPLEVPLNNTSMNATEDHPETNAVD 1156  
QY 112 LGCALDTSPTGQHPFADPTPIGASWSISPTTGGCLPWGWTQTPPTVPEDKQPGEDL 170  
DB 1157 VPKYI---PGLEDEVQPNLIP-----RRPSYEKTRMQRQDEL 1192

RESULT 2  
T04859  
extensin homolog F28A21.80 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999  
C/Accession: T04859  
R:Bevan, M.; Mueller, M.W.; Muendel, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; May  
Submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15387  
A/Accession: T04859  
A:Molecule type: DNA







Db 142 CKPCGCGPPGP-PGPPGAPGDPGEAGCTGREGTDAAPSGPPGPPGAGAGAPG-PAG 199

Qy 101 SEDHGHGRATQLPGALDTSSTGTOHPFAPETP-IPGASWSISPTTGGLLPWGWTGTPPV 159

Db 200 E-----PG-----TPAISEPLTPGAFGEPSGSPGPPGAPG- GNDGPPGP- 241

Qy 160 PEDKGPGEEDLRHPGTSLPADLPGRPPPCHPPTGT 194

Db 242 PGPKGAPGP-DGPPG-----ADGSGSPGPPGPPGAGT 271

## RESULT 10

T19288  
hypothetical protein C15A11.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T19288

R:Gardner, A.

Submitted to the EMBL Data Library, September 1996

Accession: T19103

Status: preliminary; translated from GB/EMBL/DBJ

Molecule type: DNA

A:Residues: 1-316 <M11>

A:Cross-references: EMBL:279694; PIDN:CA801958.1; GSPDB:GN00019; CESP:C15A11.6

A:Experimental source: clone C15A11

C:Genetics:

A:Gene: CESP:C15A11.6

A:Map position: 1

A:Introns: 47/3

C:Superfamily: unassigned collagens

Query Match 12.5%; Score 150.5; DB 2; Length 316;  
Best Local Similarity 30.9%; Pred. No. 0.0054;  
Matches 56; Conservative 13; Mismatches 73; Indels 39; Gaps 12;

Qy 44 PGYH-RLLCCVSCIKKASGPNPGR-DGPTSLPFPSTPGSPL-----PGQDPADAAP 94

Db 125 PSHHQKQECIKCPQGLPGAPGAPGQGPGRG--PNGNPGAPAGGCGGQPPGPPGASAGS 182

Qy 95 PGKPE-----TGSSEHGHGRATQLPGALDTSSTGTOHPFAPETP-IPGASWSISPTTGGT 148

Db 183 PGQAGACAGNCGSPKSGQKRGRLPG--PSCAPGSGGP--PGAGGQSGSNAAPGAPGPG- 237

Qy 149 PWGWTGTPPVPEDEKGPGEEDLRHPGTSLPADLPGRPPPCHPPTGTTTLCVLRDSQLRG 208

Db 238 PAG-----PNGQPG-----HPGQD---GQPGAPGNDGTPGSDAAYCPPTRSSVLR 280

Qy 209 H 209

Db 281 H 281

## RESULT 11

T21314  
hypothetical protein F23H12.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T21314

R:Kershaw, J.

Submitted to the EMBL Data Library, June 1996

A:Reference number: Z19405

A:Accession: T21314

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-301 <M11>

A:Cross-references: EMBL:274472; PIDN:CAA98942.1; GSPDB:GN00023; CESP:F23H12.4

A:Experimental source: clone F23H12

C:Genetics:

A:Gene: CESP:F23H12.4

A:Map position: 5

A:Introns: 54/3; 285/2  
C:Superfamily: unassigned collagens

Query Match 12.4%; Score 149.5; DB 2; Length 301;  
Best Local Similarity 34.9%; Pred. No. 0.006;  
Matches 59; Conservative 6; Mismatches 69; Indels 35; Gaps 13;

Qy 51 CVSCIKKASGPNPGR-----DGPPTSLP-----PSPGSPLP--GPODPADAAPPKPTG 100

Db 147 CKPCGCGPPGP-PGPPGAPDGPAGCTGRRGTDAAPSGPPGPPGAGAGAPG-PAG 204

Qy 101 SEDHGHGRATQLPGALDTSSTGTOHPFAPETP-IPGASWSISPTTGGLLPWGWTGTPPV 159

Db 205 E-----PG-----TPAISEPLTPGAFGEPSGSPGPPGAPG- GNDGPPGP- 246

Qy 160 PEDKGPGEEDLRHPGTSLPADLPGRPPPCHPPTGTTTLCVLRDSQLRG 208

Db 247 PGPKGAPGP-DGPPGVDGSGSPG-PGAPGTPGEGKIC--PKYCALDG 291

## RESULT 12

T32248  
hypothetical protein T15B7.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T32248

R:Pauley, A.; Gattung, S.

Submitted to the EMBL Data Library, September 1997

A:Description: The sequence of C. elegans cosmid T15B7.

A:Reference number: Z21139

A:Accession: T32248

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-325 <PAU>

A:Cross-references: EMBL:AF022985; PIDN:AA69961.1; GSPDB:GN00023; CESP:T15B7.4

A:Experimental source: strain Bristol N2; clone T15B7

C:Genetics:

A:Gene: CESP:T15B7.4

A:Map position: 5

A:Introns: 266/1

C:Superfamily: unassigned collagens

Query Match 12.4%; Score 149; DB 2; Length 325;  
Best Local Similarity 28.3%; Pred. No. 0.0069;  
Matches 52; Conservative 10; Mismatches 62; Indels 60; Gaps 9;

Qy 44 PGYHRLCCVSCIKKASGPNPGRDGPPTSLPFPSTPGSPL-----PGQDPADAAP 96

Db 128 PGF---CITCPAGAPGP-AGPPGAPGPKGNNGGATGATGAGSGRGPPGPPGAGGASPG 182

Qy 97 KPTGSEHGHGRATQLPGALDTSSTGTOHPFAPETP-----IPGASWSISPTTGG 147

Db 183 QP-----GHPGSPGPNRGGRGRLPSPGPPGPPG 217

Qy 148 LPW-----GWTGTPP--VPEDKGPGEEDLRHPGTSLPADLPGRPPPCHPPTGTTTLCV 200

Db 218 GPGQPGHSGAGSPGPGGPPGSPGPG---HSGND---GVPGAGNCGSPGSDAATPC 270

Qy 201 PRDS 204

Db 271 PARS 274

## RESULT 13

T18262  
S-layer protein - Clostridium thermocellum

C:Species: Clostridium thermocellum

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T18262

R:Fujino, T.; Beguin, P.; Aubert, J.P.

J. Bacteriol. 175, 1891-1899, 1993

e.  
 A:Title: Organization of a *Clostridium thermocellum* gene cluster encoding the cellulose  
 A:Reference number: Z18847; MUID:93209931  
 A:Accession: T18262  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1664 <EU>  
 A:Cross-references: EMBL:X67506; NID:9296879; PTD:9296881; PIDN:CAA47841.1

Query Match	12.3%;	Score 148;	DB 2;	Length 1664;
Best Local Similarity	26.8%;	Pred. No. 0.035;		
Matches	55;	Conservative	14;	Mismatches 78;
			Indels	58;
			Gaps	11

1 PEGQWVQSSQHPHPI-NKISTPEPCGRSVFCQMEVDRCSIPYGRILQCVSCIKRAS 59  
802 PSDEPPTSETPPEEPIPIPTDPSDEPTPSDEPT-----PSD 835  
QY 60 GPNPGDPGPGTSLP-PFSTPGSELP-PGQD---PADAAEPGKPTGSEDHQGRATOLP 113  
836 EPPTPSDEPTPSDEPTSETPPEEPIPIPTDPSDEPTPSDEPTPSDEPTPSDE-----P 886  
DQ 114 GADTSSPGQHPFAETPIPGASMSIST-----TPGGLFWGWTQTPIPYVEDKGGQGE 168  
DQ 887 TPSEPTPSE---TPREEPIPTDPSDEPTPSDEPTPSDE---TPSEDEPTPSDEPTPSE 939  
QY 169 DLRHPTSLPADLGRPPPC-NPT 192  
DQ 940 TPPEP---TPPTDPSDEPTPSDEPT 961

RESULT 14  
T24590  
hypothetical protein T06E4.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C:Accession: T24590  
R:Lloyd, C.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19910  
A:Accession: T24590  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-290 <MIL>  
A:Cross-references: EMBL:Z70756; PIDN:CAA94792.1; GSPDB:GN00023; CESP:T06E4.6  
A:Experimental source: clone T06E4  
          genetics:  
          gene: CESP:T06E4.6  
          map position: 5  
          introns: 46/3  
          superfamily: unassigned collagens

[illegible]

RESULT 15  
T20605  
hypothetical protein F08G5.4 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C:Accession: T20605  
R:Steward, C.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19299  
A:Accession: T20605  
A:Status: preliminary; translated from GB/EMBL/DBD1  
A:Molecule type: DNA  
A:Residues: 1-299 <WILL>  
A:Cross-references: EMBL:Z70682; PIDN:CAA94581.1; GSPDB:GN00022; CESP:F08G5.4  
A:Experimental source: clone F08G5  
C:Genetics:  
A:Gene: CESP:F08G5.4  
A:Map position: 4  
A:Introns: 47/3  
C:Superfamily: unassigned collagens

Query Match	12.1%;	Score 145.5;	DB 2;	Length 299;
Best Local Similarity	35.4%;	Pred. No. 0.011;		
Matches 51;	Conservative 7;	Mismatches 51;	Indels 35;	Gaps 11;

OY	51	CVCSCIKASGPNMGD-----PGTSLP----	-PESTNGSPL-	GPD-----	-PADAAE--	94
Db	136	CICPCPAGEAGP-AGFDGAFGAPFDDQGPQGDPAGCGPAGFPAGPPGDCAGAFGGDGAFGA	19			
OY	95	ECKPTSEDHONGKATOLPCALDTISSTCQHPRAFE-----TPIGASKNISPTTPG	146			
Db	195	EGAP-GGDGQKG--TGLEPA-----PGADPGQPSGNPDGDAAGAPQAQAPGAPGPD	245			
OY	147	GLPWMTOTPTVPPE-DKGGDEED	169			
Db	246	GQPGQAGQDGEAGPBGNAAGQGCAD	269			

RESULT 16  
T19291  
hypothetical protein C15A11.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C:Accession: T19291  
R:gardner, A.  
Submitted to the EMBL Data Library, September 1996  
A:Reference number: Z19103  
A:Accession: T19291  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-316 <W1>  
A:Cross-references: EMBL:Z79694; PIDD:CA01961.1; GSPDB:GN00019; CESP:C15A11.5  
C:Genetics:  
A:Gene: CESP:C15A11.5  
A:Map position: 1  
A:Introns: 4//3  
A:Superfamily: unassigned collagens

Query Match	12.1%	Score 145.5	DB 2	length 316
Best Local Similarity	30.9%	Pred. No. 0.011		
Matches	56	Conservative	12	Mismatches 74; Indels 39; Gaps 12.
QY	44	PGYH-RLCCVSCIKAKSGPNPGR--DPCGTSLLPFPSTGSPV-----PGQDDPAALAEPR	94	
DB	125	PSHHQKQECIKCPQGLPGRPACVPQGPCKG--PKNGNCAIAGGGGGCPRPFPDASADS	182	
QY	95	PGKR-----TSGEDHQHRATQLPQALDTSPPGTHPFAETPR-IPQASWISISPTTGGGL	148	
DB	183	PGQAGAPGNPGSPGKSGRGRGLPG--PGCAPCPQGG--PGAGGPGSGNAPGAPAPPG-	237	
QY	149	PGSGWQTPTPVPEDKGGQGGEDLRHPRGSLPADLRGRPEPECHPTGTFFITLCVLPKPSQLRG	208	
DB	238	PAG-----PNGQPG---HPGQD---GGGAGAGNDGTGSDAAYCPCTRTSYLR	280	

QY 209 H 209  
Db 281 H 281

## RESULT 17

T05717

probable extensin - barley (fragment)

C:Species: Hordeum vulgare (barley)

C&gt;Date: 09-Apr-1999 #sequence\_revision 09-Apr-1999 #text\_change 20-Jun-2000

C:Accession: T05717

R:Doan, D.N.P.; Sturaro, M.; Olsen, O.A.

Submitted to the EMBL Data Library, July 1997

A:Description: Characterization of a nuclear cDNA encoding a probable extensin from de

A:Reference number: Z15429

A:Accession: T05717

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1330 &lt;DOA&gt;

A:Cross-references: EMBL:Z98204; PIDN:CAB10894.1

A:Genetics:

A:Gene: ex1

A:Map position: 2

C:Superfamily: hydroxyproline-rich glycoprotein

Query Match 12.1%; Score 145.5; DB 2; Length 330;  
Best Local Similarity 36.3%; Pred. No. 0.012;  
Matches 49; Conservative 3; Mismatches 70; Indels 13; Gaps 7;

QY 61 PNRPRDGPSTLPRFSGSLRGPDRADAEPRGKPTGSEDHQGRATQLPGALDTSS 120

Db 203 PKESP-PAKRPTRPPYKPPTRPAQKRPRTTPRAKRPRTTPRAKRPRTTPRAKRPRT 261

QY 121 PGTGHPAREPTPLPGASMSISPTTGGIPMGWTQTPRPVEDKRGQPEDLRN-PGTSLP 178

Db 262 P-TPRAKRPRTTPRAH--KPTTP--TPAVKRPRTTPR-PADKRPRTTPRAKRPRTTPR 314

QY 179 ---ADLRGPRPPCH 190

Db 315 AYKAPRPSPPPPYH 329

## RESULT 18

T31065

diaphanous protein homolog p140mdia - mouse

C:Species: Mus musculus (house mouse)

C&gt;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T31065

R:Watanabe, N.; Madanle, P.; Reid, T.; Ishizaki, T.; Watanabe, G.; Kakizuka, A.; Saito,

J. 16, 3044-3056, 1997

Title: P140mdia, a mammalian homolog of Drosophila diaphanous, is a target protein for

reference number: Z20961; MUID:97357293

A:Accession: T31065

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1125 &lt;WAT&gt;

A:Cross-references: EMBL:U96963; NID:g2114472; PID:g2114473; PIDN:AAC53280.1

A&gt;Note: binds to GTP-bound form of Rho and binds to profilin

Query Match 12.1%; Score 145.5; DB 2; Length 1255;  
Best Local Similarity 30.0%; Pred. No. 0.039;  
Matches 66; Conservative 7; Mismatches 80; Indels 67; Gaps 11;

QY 3 GQWVPSGLPLPINKISSTPCTGDRSVFCQMEVLDRC---SIFGYHRLCCVSCIKK 57

Db 596 GGVVPPSPPLPPGICIPPLPLPGCA-----CIPPPQLRG-----S 632

QY 58 ASGRPPGDPPTSLPSTSTGSLPG-----PQDPAADAEPRGKPTGSEDHQGRAT 110

Db 633 AAIPPPPLPGVASIP---PPPLPGATATIPPPPLPGATATIPPPP----- 676

QY 111 QLPGALDTSSPGTQHPAPAREPTPLPGASMSISPTTGGIPMGWTQTPRPVEDKRGQPEDL 170

Db 677 -LPG-----GTGTP-PPPPPLPG---SVCVPPPPPLPG-GGGLPPPPPPPPGAGPIRP 723

## RESULT 19

A39808

proteoglycan core protein, cartilage - bovine (fragments)

N:Alternate names: aggrecan; aggregating cartilage proteoglycan

C:Species: Bos primigenius taurus (cattle)

C&gt;Date: 20-Mar-1992 #sequence\_revision 23-Mar-1995 #text\_change 13-Aug-1999

C:Accession: A34234; A27752; A39808; A27751; E29164; B27751; D27751; E27751;

R:Antonsson, P.; Helnegard, D.; Oldberg, A.

J. Biol. Chem. 264, 16170-16173, 1989

A:Title: The keratan sulfate-enriched region of bovine cartilage proteoglycan consist

A:Reference number: A34234; MUID:89380219

A:Accession: A34234

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 128-621 &lt;ANT&gt;

A:Cross-references: GB:J05028

R:Oldberg, A.; Antonsson, P.; Helnegard, D.

Biochem. J. 243, 255-259, 1987

A:Title: The partial amino acid sequence of bovine cartilage proteoglycan, deduced fro

A:Reference number: A27752; MUID:87270630

A:Accession: A27752

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 622-1340 &lt;OLD&gt;

R:Sandy, J.D.; Boynton, R.E.; Flannery, C.R.

J. Biol. Chem. 266, 8198-8205, 1991

A:Title: Analysis of the catabolism of aggrecan in cartilage explants by quantitation

A:Reference number: A39808; MUID:91217051

A:Accession: A39808

A&gt;Status: preliminary

A:Molecule type: protein

A:Residues: 1-28;59-82;131-137,'QSEF',142-149;196-207;226-249;1137-1143;1252-1267;127

R:Perin, J.P.; Bonnel, F.; Jolles, P.

FEBS Lett. 206, 73-77, 1986

A:Title: Structural relationship between link proteins and proteoglycan monomers.

A:Reference number: A27751; MUID:87005253

A:Accession: A27751

A:Molecule type: protein

A:Residues: 29-58;74-130;174-175,'A',177-204;208-225 &lt;PER&gt;

R:Perin, J.P.; Bonnel, F.; Jolles, P.

FEBS Lett. 176, 37-42, 1984

A:Title: Sequence data concerning the protein core of the cartilage proteoglycan mono

A:Reference number: A91327; MUID:85027710

A:Accession: E29164

A:Molecule type: protein

A:Residues: 1230-1249 &lt;PE2&gt;

C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;

C:Keywords: glycoprotein

F:8-28/Domain: link protein repeat homology (fragment) &lt;LNK1&gt;

F:29-58/Domain: link protein repeat homology (fragment) &lt;LNK2&gt;

F:80-146/Domain: link protein repeat homology (fragment) &lt;LNK3&gt;

F:157-248/Domain: link protein repeat homology (fragment) &lt;LNK4&gt;

F:1130-1250/Domain: C-type lectin homology &lt;LCH&gt;

F:1257-1313/Domain: complement factor H repeat homology &lt;PHD&gt;

Query Match 12.1%; Score 145.5; DB 2; Length 1340;  
Best Local Similarity 27.0%; Pred. No. 0.042; 71; Indels 75; Gaps 17;  
Matches 64; Conservative 27; Mismatches 71; Indels 75; Gaps 17;

QY 5 WVPQSGPLPINKISSTPCTGD---RSVFC---QMEVLDRCISIPGYHRLCC--VSCI 55

Db 200 WLADGSLRPI-VTPRACGGDKGVRTVLYLPQTGLD---PLSRHACGFGVSA 254

```

QY 56 ----KKASQPNNGPD-----RG-----PISLPPFS-----TP 78
      ::::: ||| |||
Db 255 PPSPEEGSAPTRAGPVEEEMMYTGVGVAAVPIGGEITTAIGFTVEENKTEWELATP 314
      ::::: ||| |||
QY 79 GSPILPGPODPADAAEPCKPPIG--SEDHGQ-RATQIPGALDITSSGCHQHPAETPLPG 135
      ::::: ||| |||
Db 315 AGTLPLPGIP-----PTMPPTGGEATEHTGSPKATVEPASEKPPV-SSEPPPEEPPV- 367
      ::::: ||| |||
QY 136 ASMSISPTTPGGLPMGWOTPLPYVEDKQGPGEGLDRHGSNPLADLPGPPECPHT 192
      ::::: ||| |||
Db 368 ---SEKPPPEEL---PPSKPPPSKPPPSSE-----PPPSKPP-PPPEELPPS 410

```

SSULT 20  
 19630  
 Species: Bos primigenius taurus (cattle)  
 C.Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 04-Mar-2000  
 C.Accession: T42630  
 R.Hering, T.M.Kollar, J.; Huynh, T.D.  
 submitted to the EMBL Data Library, September 1996  
 A.Description: Complete coding sequence of bovine aggrecan: comparative structural analysis  
 A.Reference number: Z22182  
 A.Accession: T42630  
 A.Status: preliminary; translated from GB/EMBL/DBDJ  
 A.Molecule type: mRNA  
 A.Residues: 1-2327 <HER>  
 A.Cross-references: EMBL:U76615; NTD:G1730259; P1D:G1740260; P1DN:ABB38524.1  
 A.Experimental source: articular chondrocytes  
 C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; ECG  
 C:Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein

Query Match	12.18;	Score 145.5;	DB 2;	Length 2327;
Best Local Similarity	27.08;	Pred. No. 0.069;		
Matches 64;	Conservative 27;	Mismatches 71;	Indels 75;	Gaps 17.

QY 56 ----KKASGPNPCPD-----DG-----P<sup>11</sup>SLPPFS-----TP 78  
:: | | | | :: | | | |  
Db 690 PPSPEEESGSPAGADPVDEEMNTQVGGVAAPVIGEEFTTALPGFVPEPNKTEFWLAVTD 749

```

13  GSEYFGGPDADAAEPKPKPG--SEBHQHG-RATQLPGALDTSSTGTHPFAEPPIPG 135
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
750 AGTLPLPGIP-----PTWPPGEATEEHTEGSATVPSASEKPP-SEEPFPPPEEPF- 802

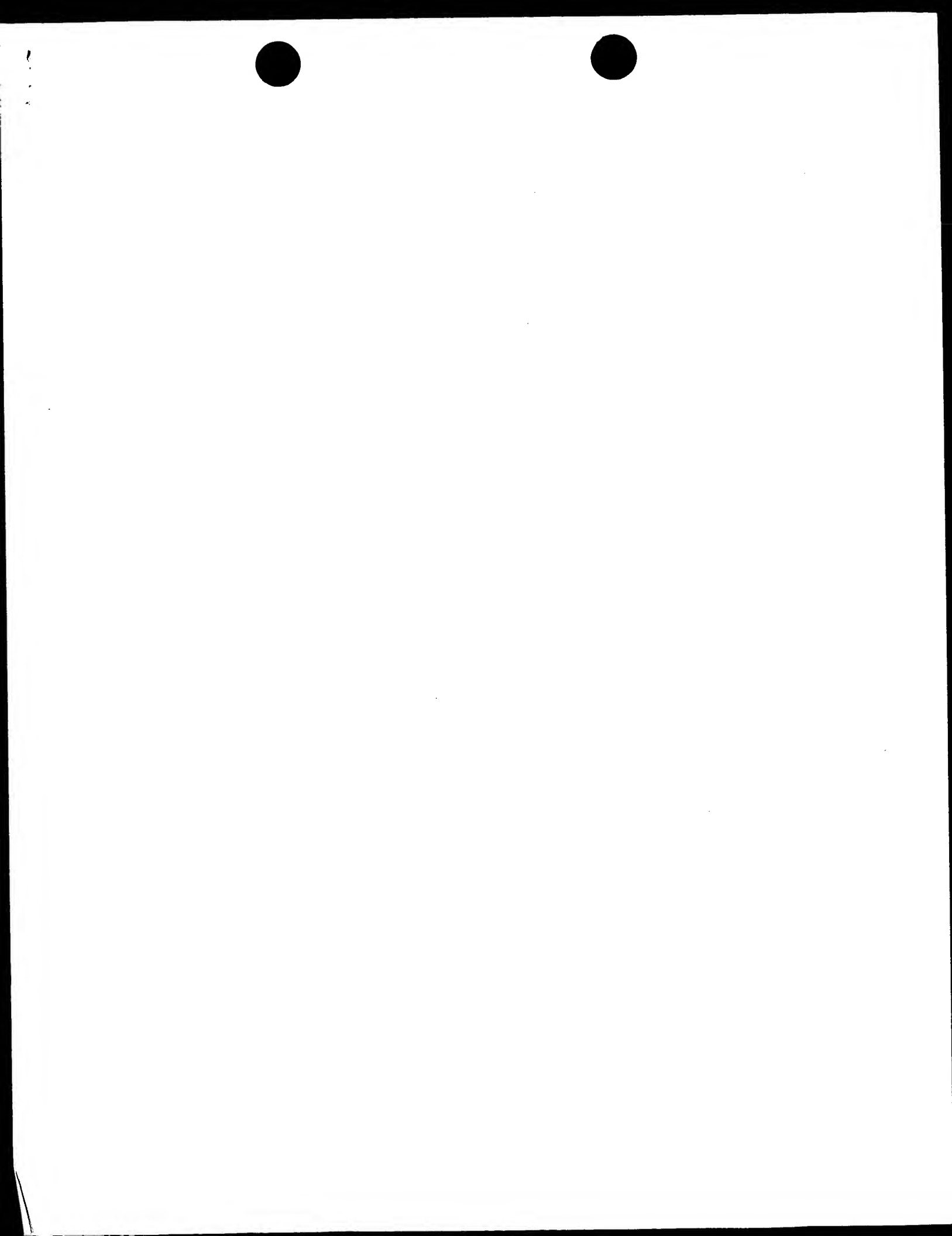
```

```

27      ASMSISITIPGGLWGWITQTPVPVEDKQPGEDLRHPTSLPADLRGRPPPCHPPT 192
      | | | : | | | : | | | :
130  ---SEKFPPEEL----FPSEKFPSEKFPSE-----PFPSEKP-FPPEELFPS 845
      | | | : | | | : | | | :
Db    803  ---SEKFPPEEL----FPSEKFPSEKFPSE-----PFPSEKP-FPPEELFPS 845

```

Search completed: August 29, 2002, 00:09:27  
Job time: 408 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 00:06:29 ; Search time 37.54 Seconds

(Without alignments)  
216.598 Million cell updates/sec

Title: US-09-938-330-22\_COPY\_1040\_1249

Sequence: 1 PEGQWVPSGPIHPINKISS.....PTGFTLCVLPDSQLNGHT 210

ing table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	211.5	17.5	1205 1	AT52_BOVIN
2	207.5	17.2	1211 1	AT53_HUMAN
3	204	16.9	1201 1	AT53_HUMAN
4	155.5	12.9	317 1	Y035_CAEL
5	151.5	12.6	210 1	CAC2_HAEC
6	151	12.5	296 1	CC01_CAEL
7	148.5	12.3	555 1	GLP1_CHRE
8	148	12.3	1664 1	GLP1_CHRE
9	145.5	12.1	1255 1	DIA1_MOUSE
10	145.5	12.1	2364 1	DIA1_MOUSE
11	145	12.0	331 1	PRP1_HUMAN
12	145	12.0	1466 1	PRP1_HUMAN
13	142	11.8	295 1	CAC3_HAEC
14	141.5	11.7	316 1	CC12_CAEL
15	141.5	11.7	316 1	CC13_CAEL
16	140.5	11.7	266 1	YXWK_CAEL
17	140.5	11.7	1690 1	CC44_HUMAN
18	139	11.5	298 1	CC34_CAEL
19	138.5	11.5	302 1	CCPC_CAEL
20	138.5	11.5	1183 1	DRPL_RAT
21	138.5	11.5	1464 1	CA13_MOUSE
22	138.5	11.5	1496 1	CA25_HUMAN
23	137.5	11.4	503 1	WAIP_HUMAN
24	137	11.4	1248 1	DIA1_HUMAN
25	136	11.3	251 1	PRP2_HUMAN
26	136	11.3	1459 1	PRP2_MOUSE
27	135.5	11.2	754 1	CA54_CANF
28	135.5	11.2	1097 1	S24C_ARATH
29	135	11.2	301 1	CC02_CAEL
30	134.5	11.2	279 1	Y091_NPYOP
31	134.5	11.2	1464 1	CA11_HUMAN
32	134	11.1	2944 1	CA17_HUMAN
33	134	11.1	1	CA17_HUMAN

## Description

P79331 b adamts-2  
O95450 h adamts-2  
O15072 homo sapien  
Q09456 caenorhabdi  
P16252 haemochus  
P08124 caenorhabdi  
O91P06 chlamydomon  
Q06852 cistridium  
P13608 bos taurus  
P04280 mus musculu  
P16253 haemochus  
P20630 caenorhabdi  
P20631 caenorhabdi  
Q21184 caenorhabdi  
P34687 caenorhabdi  
P17625 caenorhabdi  
P51256 ractus norv  
P08121 mus musculu  
P05997 homo sapien  
O43516 homo sapien  
O03692 homo sapien  
O60610 homo sapien  
P02812 homo sapien  
P28481 mus musculu  
Q28247 canis famli  
O09291 arabidopsis  
P17656 caenorhabdi  
O10341 oryza pseu  
P02453 homo sapien  
Q02368 homo sapien

## ALIGNMENTS

RESULT 1	ID	AT52_BOVIN	STANDARD:	PRT: 1205 AA.
AC	P79331	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)			
DE	(Procollagen I/II amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI).			
GN	ADAMTS2 OR NPI.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovine; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=SKIN;			
RX	MEDLINE=9725960; PubMed=9122202;			
RA	Colige A., Li S.W., Steron A.L., Nusgens B.V., Prockop D.J.,			
RA	Lapierre C.M.;			
RT	"CDNA cloning and expression of bovine procollagen I N-proteinase: a new member of the superfamily of zinc-metalloproteinases with binding sites for cells and other matrix components.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).			
RL	[2]			
RN	PARTIAL SEQUENCE.			
RX	MEDLINE=95348096; PubMed=7622483;			
RA	Colige A., Beschlin A., Samyn B., Goebels Y., Van Beeumen J.,			
RA	Nusgens B.V., Lapierre C.M.;			
RT	"Characterization and partial amino acid sequencing of a 107-kDa procollagen I N-proteinase purified by affinity chromatography on immobilized type XIV collagen.";			
RT	J. Biol. Chem. 270:16724-16730(1995).			
RL	- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN COLLAGEN BIOSYNTHESIS.			
CC	- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain alpha-1(I) at Pro-1-Gln and of alpha-1(I) and alpha-2(I) chains at Ala-1-Gln.			
CC	- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).			
CC	- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO COLLAGEN TYPE XIV.			
CC	- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).			
CC	- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE mRNA LEVELS WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.			
CC	- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-I DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.			
CC	- PTM: THE N-TERMINUS IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY			
CC	- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY			

CC SIMILARITY).

CC -1- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOSPAPRAXIS, A

CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN

CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN

CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.

CC

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL: X66389; CAA65253.1; -

CC InterPro: IPR001762; Disintegrin.

CC InterPro: IPR002870; Pep\_M12B\_propep.

CC InterPro: IPR001590; Repolysin.

CC InterPro: IPR000884; TSP1.

CC InterPro: IPR000130; Zn\_MTPeptide.

CC Pfam: PR01562; Rep\_M12B\_propep; 1.

CC Pfam: PR01421; Repolysin; 1.

CC SMART: SM00209; TSP\_1; 4.

CC DR PROSITE: PS50215; ADAM\_MEPRO; 1.

CC DR PROSITE: PS50092; TSP1; 1.

CC DR PROSITE: PS00142; ZINC\_PROTEASE; FALSE\_NEG.

CC DR PROSITE: PS00427; DISINTEGRIN\_1; FALSE\_NEG.

CC KM Hydrolysis: Metalloprotease; Zinc; Signal: Glycoprotein; Zymogen;

CC Repeat: Collagen degradation; Extracellular matrix; Heparin-binding.

CC FT SIGNAL 1 28

CC FT PROPEP 29 253

CC FT CHAIN 254 1205

CC FT METAL 402 402

CC FT ACT\_SITE 403 403

CC FT METAL 406 406

CC FT METAL 412 412

CC FT METAL 417 417

CC FT METAL 474 554

CC FT METAL 555 611

CC FT METAL 612 716

CC FT METAL 717 845

CC FT METAL 846 905

CC FT METAL 906 968

CC FT METAL 969 1024

CC FT METAL 685 687

CC FT SITE 31 35

CC FT DOMAIN 177 180

CC FT DOMAIN 104 104

CC FT CARBOHYD 245 245

CC FT CARBOHYD 942 942

CC FT CARBOHYD 943 943

CC FT CARBOHYD 987 987

CC FT CARBOHYD 1025 1025

CC FT CARBOHYD 1092 1092

CC FT CARBOHYD 1139 1139

CC FT CARBOHYD 1144 1144

CC SEQUENCE 1205 AA; 133887 MW; 785B232A45320371 CMC64;

Query Match 17.5%; Score 211.5; DB 1; Length 1205;

Best Local Similarity 29.1%; Pred. No. 1.2e-05;

Matches 52; Conservative 24; Mismatches 68; Indels 35; Gaps 4;

OY 4 QWVPSGGLPIPKISSTPECTGDSVFCQMEVLDRTCSIGYRLCCVSC----- 54

DB 1037 QWLSRPDPSNVSQVETSRGCGDKSVFCRMEVLSRYCSIGYRLCCVSCPNHDLTDV 1096

OY 55 IKKAGPRNGDPRGTSLSPPSTGSPSPL---PGQDPADAAEPFGKPTGSGDHQGRATQ 111

DB 1097 DDRAPPSGKNDIEELMPTLSVPTLVMEWQPPGICPLEVPLNTSSTATNEDHETNAVD 1156

OY 112 LPGAIDSSPQTHPPAPETIPPCANSISPTTGGILPMGHTQTPPVPPDKGPGDGL 170

DB 1157 VPKYI---PGLDEVPQPNLIP-----RRPSYEXTRNRIQEL 1192

RESULT 2

AT22\_HUMAN STANDARD; PRT; 1211 AA.

ID AT22\_HUMAN

AC 095450;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and

DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS2)

DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I

DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI)

DE (Procollagen I/II amino-propeptide processing enzyme).

GN ADAMTS2 OR PCNP1 OR PCPNI.

OS Homo sapiens (Human).

OC Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS LPNP1 AND SPNP1).

RC TISSUE-Skin;

RA MEDLINE=99347935; PubMed=10417273;

RA Collige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,

RA Werleicki W., Wilcox W., Krakow D.J., Cohn D.H., Reardon W.,

RA Byers P.H., Lapiere C.M., Prockop D.J., Nussgens B.V.,

RA "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis

RA are caused by mutations in the procollagen I N-proteinase gene.";

RA Am. J. Hum. Genet. 65:308-317(1999).

RT

RL

CC -1- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR

CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO

CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN

CC COLLAGEN BIOSYNTHESIS.

CC -1- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain

CC alpha-1(I) at Pro-1-Gln and of alpha-1(I) and alpha-2(I) chains

CC at Ala-1-Gln.

CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -1- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO

CC COLLAGEN TYPE XIV (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR

CC MATRIX (BY SIMILARITY).

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNP1 (SHOWN HERE) AND SPNP1;

CC ARE PRODUCED BY ALTERNATIVE SPLICING. SPNP1 HAS NO SIGNIFICANT N-

CC PROCOLLAGEN PEPTIDASE ACTIVITY.

CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON

CC AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.

CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT

CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY

CC SIMILARITY).

CC -1- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF EHLERS-DANLOS

CC SYNDROME TYPE VIIC (EDS-VIIC). A RECESSIVELY INHERITED DISORDER

CC CHARACTERIZED CLINICALLY BY SEVERE SKIN FRAGILITY AND

CC BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN INCOMPLETELY

CC PROCESSED AT THE AMINO TERMINUS.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -1- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.

CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.

CC

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AJ003125; CAA05880.1; -  
 DR MEROPS: M12.301; -  
 DR MIM: 604539; -  
 DR MIM: 225410; -  
 DR InterPro: IPR001762; Disintegrin.  
 DR InterPro: IPR002870; Pep\_M12B-propep.  
 DR InterPro: IPR001590; Repolysin.  
 DR InterPro: IPR000884; TSP1.  
 DR InterPro: IPR00130; Zn\_MTPeptide.  
 DR Pfam: PF01562; Pep\_M12B-propep; 1.  
 DR Pfam: PF01421; Repolysin; 1.  
 DR Pfam: PF00090; TSP1; 4.  
 DR SMART: SM00209; TSP1; 4.  
 DR PROSITE: PS50215; ADAM\_MEPRO; 1.  
 DR PROSITE: PS50092; TSP1; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; FALSE NEG.  
 DR PROSITE: PS00427; DISINTEGRIN\_L1; FALSE NEG.  
 DR Hydrolase: Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;  
 KW Alternative splicing.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT PROPEP 30 253 BY SIMILARITY.  
 FT CHAIN 254 1211 ADAMTS-2.  
 FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 409 409 BY SIMILARITY.  
 FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DOMAIN 480 560 DISINTEGRIN-LIKE.  
 FT DOMAIN 561 617 TSP TYPE-1 1.  
 FT DOMAIN 618 722 CYS-RICH.  
 FT SITE 691 693 CELL ATTACHMENT SITE (POTENTIAL).  
 FT DOMAIN 723 851 SPACER.  
 FT DOMAIN 852 911 TSP TYPE-1 2.  
 FT DOMAIN 912 974 TSP TYPE-1 3.  
 FT DOMAIN 975 1030 TSP TYPE-1 4.  
 FT DOMAIN 40 43 POLY-ALA.  
 FT DOMAIN 185 188 POLY-GLU.  
 FT CARBOHYD 112 112 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 251 251 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 949 949 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 993 993 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 1031 1031 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 1098 1098 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 1145 1145 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 1150 1150 N-LINKED (GLCNAC) (POTENTIAL).  
 FT VARSPLIC 544 566 HCFGHCIMLPDLKRDGSMGA -> FRGGAIAHCYPS  
 TILGCGRMA (IN ISOFORM SPNFI).  
 VARSPLIC 567 1211 MISSING (IN ISOFORM SPNFI).  
 SEQUENCE 1211 AA; 134722 MW; BECEEF25C23CAD2D CRC64;

Query Match 17.2%; Score 207.5; DB 1; Length 1211;  
 Best Local Similarity 34.7%; Pred. No. 2, 1e-05;  
 Matches 50; Conservative 14; Mismatches 67; Indels 13; Gaps 3;

QY 4 QWVPSGPHLPINKSSTEPCTGDSVFCQMEVLDRCYSPGYHRLCYSC-----I 55  
 DB 1043 QWLSRPDSDPSIRKISGHCQKSIKRMELVSRCSISGYKLSCKSNLYNNLTNV 1102  
 QY 56 KASGPNPDP-----DPGTSI.LPPESTPGSLPQGDADAAEPGKKTGSEDOHGRATO 111  
 DB 1103 EKRIPPPGKHNDIVFMPITLVPVVAEVRPSPSTPLEVPLNASTNATEDHPETNAVYD 1162  
 QY 112 LPGAIDTSSPGTQHP-FAPETPIP 134  
 DB 1163 EBYKIHGLEDEVQPNLIPRRSP 1186

RESULT 3  
 ID ATSS3\_HUMAN STANDARD; PRT; 1201 AA.  
 AC 015072;  
 DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE ADAMTS-3 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 3) (ADAM-TS 3) (ADAM-TS3) (Fragment).  
 GN ADAMTS3 OR KIA0366.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.;  
 RL DNA Res. 4:141-150(1997).  
 CC -1- CORFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
 CC MATRIX (BY SIMILARITY).  
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
 CC -1- PFM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.  
 CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB002364; BAA20821.1; -  
 DR MEROPS: M12.220; -  
 DR MIM: 605011; -  
 DR InterPro: IPR001762; Disintegrin.  
 DR InterPro: IPR002870; Pep\_M12B-propep.  
 DR InterPro: IPR001590; Repolysin.  
 DR InterPro: IPR000884; TSP1.  
 DR Pfam: PF01562; Pep\_M12B-propep; 1.  
 DR Pfam: PF01421; Repolysin; 1.  
 DR Pfam: PF00090; TSP1; 4.  
 DR SMART: SM00209; TSP1; 4.  
 DR PROSITE: PS50215; ADAM\_MEPRO; 1.  
 DR PROSITE: PS50092; TSP1; 2.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; FALSE NEG.  
 DR PROSITE: PS00427; DISINTEGRIN\_L1; FALSE NEG.  
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix; Heparin-binding;  
 FT SIGNAL 1 25 POTENTIAL.  
 FT PROPEP 26 245 BY SIMILARITY.  
 FT CHAIN 246 1201 ADAMTS-3.  
 FT METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 395 395 BY SIMILARITY.  
 FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 404 404 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DOMAIN 466 546 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DOMAIN 547 603 DISINTEGRIN-LIKE.  
 FT DOMAIN 604 708 TSP TYPE-1 1.  
 FT DOMAIN 709 840 CYS-RICH.  
 FT DOMAIN 841 898 SPACER.  
 FT DOMAIN 899 961 TSP TYPE-1 2.  
 FT DOMAIN 962 1013 TSP TYPE-1 3.  
 FT DOMAIN 242 245 TSP TYPE-1 4.  
 FT DOMAIN POLY-ARG.

FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 810 810 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1201 AA; 135113 MW; D54EA92BD506A3AA CRC64;

Query Match 16.9%; Score 204; DB 1; Length 1201;  
 Best Local Similarity 38.7%; Pred. No. 3.2e-05;  
 Matches 58; Conservative 12; Mismatches 36; Indels 44; Gaps 8;

QY 12 LHPINKISTEPCTGDRSVFCOMEVLDRCYSIPGYHRLCVSCIKKASG----- 60  
 DB 1008 LPPCN-----DEPCLDCKSIFCOMEVLDRCYSIPGYHRLCVSCIKKASG----- 1063  
 QY 61 -----PNEGPPPG-----PTSLPPF-----SNPG-----SPLPGQDPADAAEPGKPTG 100  
 DB 1064 THDDVISNPDDLPSLWPTSLVPHSETPAKKMSLSSISVGPNAYA-AFRPNKPDG 1122  
 DB 101 SEDHGRATOLPGALDT-----SSPGTQ 124  
 DB 1123 ANLRQ--RSAGQAGSKTVRLVTPSSPTK 1150

RESULT 4  
 ID YQ35\_CAEEL STANDARD; PRT; 317 AA.  
 AC 009456;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Putative cuticle collagen C0963.5.  
 GN C0963.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Palmer S.;

RL Submitted (NOV-1994) to the EMBL/Genbank/DDJJ databases.  
 CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).  
 CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 CC LINKS (BY SIMILARITY).  
 CC -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE  
 CC COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: Z46791; CAA86758.1; -  
 CC WormBep: C0963.5; CE01485.  
 DR InterPro: IPR002486; Collagen.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 3.  
 DR Pfam: PF01484; Col\_cuticle\_N; 1.  
 KW Hypothetical protein; Cuticle; Connective tissue; Repeat;  
 KW Multigene family; Collagen.  
 FT DOMAIN 92 124 TRIPLE-HELICAL REGION.  
 FT DOMAIN 137 199 TRIPLE-HELICAL REGION.  
 FT DOMAIN 202 264 TRIPLE-HELICAL REGION.

SQ SEQUENCE 317 AA; 31283 MW; 685DCF24612707BB CRC64;

Query Match 12.9%; Score 155.5; DB 1; Length 317;  
 Best Local Similarity 33.9%; Pred. No. 0.0065;  
 Matches 61; Conservative 9; Mismatches 63; Indels 47; Gaps 15;

QY 44 PGYHRLC-CVSCIKKASGPNPGP-DPGPTSLPFTSPGSP-----PGPDPA-DAAE 93  
 DB 123 PAHQOQECTCKCPGAGPAGPAGNPGPG--PNEGPPAHAGCGGPPGPGAGDAGS 180  
 QY 94 P-----PGKPTGSEDDHGRATOLPGALDTSSPGTQHPPAETP-IPGASWSISPTT 144  
 DB 181 PGAGAGPAGNPGRGSGRSGR---LPG--PSGRPGPGP--PGAPGPG-----SGST 227  
 QY 145 PGGLPKWTPPTPRVPRDKORQEDRLRHPTSLRALPGRRPPEPCPTGFTLCVLRPS 204  
 DB 228 PG--PAG-----PGGPPGNGPG--HpgQD---GQFAGPAGDAGPAGSMAATCPAPAS 274

RESULT 5

ID CAC2\_HAECC STANDARD; PRT; 210 AA.  
 AC P16252;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Cuticle collagen 2c (Fragment).  
 GN 2C.  
 OS Haemonchus contortus (Barber pole worm).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
 OC Trichostrongylidae; Haemonchidae; Haemonchinae; Haemonchus.  
 OX NCBI\_TaxID=6289;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=90136718; PubMed=2615789;  
 RA Shamansky L.M., Pratt D., Boissvenne R.J., Cox G.N.;

RL "Cuticle collagen genes of Haemonchus contortus and Caenorhabditis  
 RT elegans are highly conserved."  
 MOJ. Biochem. Parasitol. 37:73-86(1989).  
 CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 CC -1- MISCELLANEOUS: THIS PROTEIN SHOWS 4 POTENTIAL TRIPLE-HELICAL  
 CC REGIONS, WHICH CONTAIN GLYCINE AS EVERY THIRD AMINO ACID.  
 CC -1- MISCELLANEOUS: IN ALL NEMATODE CUTICLE COLLAGENS, THE POLYPEPTIDE  
 CC CHAINS ARE COMPLEXED WITHIN THE CUTICLE BY DISULFIDE BONDS AND  
 CC OTHER TYPES OF COVALENT CROSS-LINKS.  
 CC -1- SIMILARITY: TO CAENORHABDITIS ELEGANS CUTICLE COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: J04670; AAA29172.1; -  
 CC InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 2.  
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.  
 FT NON\_TER 1  
 SQ SEQUENCE 210 AA; 19562 MW; E15FAA9A2DF3D74B CRC64;

Query Match 12.6%; Score 151.5; DB 1; Length 210;  
 Best Local Similarity 34.0%; Pred. No. 0.0079;  
 Matches 54; Conservative 8; Mismatches 64; Indels 33; Gaps 11;

QY 51 CSCIKKASGPPGPPDPTSLPFTSPGSP-----LPGPDPAADAEPPGKPTGSED 103  
 DB 56 CKRCPGGRGP-PGP-IGPPGP--GTPGNGPAGPNDAPGPBPGRKGPGRK----- 105

QY 104 HOGRATOLPGALDTSPTGQHPFAPETP-----IPGASMSISPTTPGGLPMGWTOTPTPV 159  
 Db 106 -----AGAGCA--AGOPGAGNAPSEPLVPEPPGPPPTGPEGGPNCAP-GHPGAP-CA 155  
 QY 160 PEDKGQPEDLRHPTSLPADLPGRPPECHPTGTTLG 198  
 Db 156 PGEKGPGRGD-GHPGAPGNAGHPGPGP-GPPGEGKVC 192

RESULT 6  
 CC01\_CAEEL STANDARD; PRT: 296 AA.  
 AC P08124;  
 01-AUG-1988 (Rel. 08, Created)  
 01-AUG-1988 (Rel. 08, Last sequence update)  
 01-FEB-1996 (Rel. 33, Last annotation update)  
 Cuticle collagen 1.  
 SQT-3 OR COL-1 OR DPY-15.  
 GN Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=85105075; PubMed=2578467;  
 RA Kramer J.M., Cox G.N., Hirsch D.;  
 RT "Expression of the Caenorhabditis elegans collagen genes col-1 and  
 col-2 is developmentally regulated.";  
 RL J. Biol. Chem. 260:1945-1951(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=83050944; PubMed=7139711;  
 RA Kramer J.M., Cox G.N., Hirsch D.;  
 RT "Comparisons of the complete sequences of two collagen genes from  
 Caenorhabditis elegans.";  
 RL Cell 30:599-606(1982).

CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 LINKS.  
 CC -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE  
 COLLAGENS.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: J01047; AAA27988.1; -;  
 DR EMBL: V00147; CAA23463.1; -;  
 DR PIR: A31219; A31219.  
 DR InterPro: IPR002486; COL\_cuticle\_N.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen. 2.  
 DR Pfam: PF01484; COL\_cuticle\_N; 1.  
 DR Cuticle: Connective tissue; Repeat; Multigene family; Collagen.  
 KW DOMAIN 100 129 TRIPLE-HELICAL REGION.  
 FT CHAIN 148 174 TRIPLE-HELICAL REGION.  
 FT DOMAIN 178 204 TRIPLE-HELICAL REGION.  
 FT DOMAIN 213 278 TRIPLE-HELICAL REGION.  
 SO SEQUENCE 296 AA; 28574 MW; 843B169D2C44526A CRC64;

Query Match 12.5%; Score 151; DB 1; Length 296;  
 Best Local Similarity 36.8%; Pred. No. 0.011;

Matches 57; Conservative 3; Mismatches 59; Indels 36; Gaps 12;  
 QY 51 CVSCIKKASGPNP-----DPGTSLP-----PESTPGSLP-GPODDADAEPPGKPTG 100  
 Db 142 CKPCPGCPPP-PGPGACADPDPEAGTPCRPGTDAAGSPGRCRPPGAGEGAPG-PAG 199  
 QY 101 SEDHOGRATOLPGALDTSPTGQHPFAPETP-IPGASMSISPTTPGGLPMGWTOTPTPV 159  
 Db 200 E-----TPAISEPLTPGAPGEPODGGPPGPAPGAP-GNDGPPGP- 241  
 QY 160 PEDKGQPEDLRHPTSLPADLPGRPPECHPTGT 194  
 Db 242 PGPAGAPGD-GPPG-----ADGSGPPGPPGAGT 271

RESULT 7  
 GPL\_CHURE STANDARD; PRT: 555 AA.  
 AC G9FP06; Q03927;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Vegetative cell wall protein gpl precursor (Hydroxyproline-rich  
 glycoprotein 1).  
 GN Gpl.  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadales; Chlamydomonas.  
 OX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA PubMed=11258910;  
 RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,  
 RA Goodenough U.W.;  
 RT "Glycosylated polyproline II rods-with-kinks as a structural motif in  
 plant hydroxyproline-rich glycoproteins.";  
 RL Biochemistry 40:2978-2987(2001).

CC -1- FUNCTION: PRELIMINARY SEQUENCE FROM N.A.  
 RX MEDLINE=91017504; PubMed=1699225;  
 RA Adair W.S., Apt K.E.;  
 RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs  
 encoding cell wall hydroxyproline-rich glycoproteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).  
 CC -1- FUNCTION: Major component of the outer cell wall W6 (crystalline)  
 layer.  
 CC -1- SUBUNIT: Associates with GP2 and GP3.  
 CC -1- PTM: N-glycosylated and O-glycosylated.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF309494; AAG45420.1; -;  
 DR EMBL: M58496; AAA69706.1; ALT-seq.  
 DR GlycoSuiteDB: G9FP06; -;  
 DR InterPro: IPR003882; Pistill\_extensin.  
 DR InterPro: IPR002965; P-rich\_extensn.  
 DR PRINTS: PR01217; PRICHEXTENSIN.  
 DR PRINTS: PR01218; PSLTEXTENSIN.  
 KW Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN Gpl.  
 FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSX REPEATS.  
 FT DOMAIN 259 279 POLY-PRO.  
 FT CARBOHYD 399 399  
 FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 555 AA; 54219 MW; 6A584A9046502F5 CRC64;

Query Match 12.3%; Score 148.5; DB 1; Length 555;  
 Best Local Similarity 28.6%; Pred. No. 0.026;  
 Matches 42; Conservative 11; Mismatches 73; Indels 21; Gaps 4;

61 PNPGRDPTSLPPTSTGSG-----PLRGQDDPADAE-----PGKPGSDHQHGR 108  
 169 DIRHP---GTSIPADLPGRRPEPCHT 192  
 275 PRPPFPANTPMPSPSPSPSPAPPT 301

RESULT 8  
 CLOTM STANDARD; PRT; 1664 AA.  
 006852;  
 01-JUN-1994 (Rel. 29, Created)  
 01-JUN-1994 (Rel. 29, Last sequence update)  
 01-FEB-1996 (Rel. 33, Last annotation update)  
 Cell surface glycoprotein 1 precursor (outer layer protein B) (S-layer protein 1).  
 OLPB.  
 Clostridium thermocellum.  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 Clostridium.  
 NCBI\_TaxID=1515;  
 [1]  
 RP SEQUENCE FROM N.A.  
 STRAIN=NCIB 10682;  
 MEDLINE=9320931; PubMed=8458832;  
 Fujino T., Beguin P., Aubert J.-P.;  
 "Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein CtpA and a protein possibly involved in attachment of the cellulosome to the cell surface."  
 RT J. Bacteriol. 175:1891-1899(1993).  
 CC -1 SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.  
 CC -1 SUBCELLULAR LOCATION: Cell wall.  
 CC -1 SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC -----  
 DR EMBL: X67506; CAA47841.1; -  
 DR InterPro: IPR001119; SLH.  
 DR Pfam: PF00395; SLH; 3.  
 DR PROSITE: PS01072; SLH DOMAIN; 2.  
 KW Cell wall; S-layer; Signal; Repeat.  
 FT SIGNAL 1 28  
 FT CHAIN 29 1664  
 FT DOMAIN 36 763  
 FT REPEAT 36 191  
 FT REPEAT 207 363  
 FT REPEAT 409 565  
 FT REPEAT 607 763  
 FT REPEAT 771 1377  
 FT DOMAIN 1378 1449  
 FT DOMAIN 1453 1494  
 FT DOMAIN 1495 1565  
 FT DOMAIN 1566 1625  
 FT DOMAIN 1626 1646  
 SLH 1 (INCOMPLETE).  
 SLH 2.  
 SLH 3.  
 SLH 4 (INCOMPLETE).

SEQUENCE 1664 AA; 178194 MW; 5F396695BA9E74B CRC64;

Query Match 12.3%; Score 148; DB 1; Length 1664;  
 Best Local Similarity 26.8%; Pred. No. 0.069;  
 Matches 55; Conservative 14; Mismatches 76; Indels 58; Gaps 11;

1 PEGQWVPSGSLHPI-NKISSTPCYGRSVCQMEVLIDYCSIPGYHRLCCVSCIKKAS 59  
 802 PSDEPTSETPPEIPPTDPSDEPTPSDEPT-----PSD 835  
 60 GNPGRDPTSLP-PPSTGSLP--GPOD---PADAAEPCKPGSDHQHGRATOLP 113  
 836 EPPTSEDEPTPSDEPTSETPPEIPPTDPSDEPTPSDEPTPSDEPTPSDE-----P 886  
 114 GALTSSPGTQHFAETPIPGASWSISPTTGGLPWGTQTPTVPEDKGPGE 168  
 887 TPSPDEPTSE-----TPPEIPPTDPSDEPTPSDEPTPSDEPT-----TPSDEPTPSDEPTPSE 939  
 169 DIRHPGTSIPADLPGRRPEPC-HPT 192  
 940 TPPEP---IPTDTPSDEPTPSDEPT 961

RESULT 9  
 DIAL\_MOUSE STANDARD; PRT; 1255 AA.  
 008808;  
 15-JUL-1999 (Rel. 38, Created)  
 15-JUL-1999 (Rel. 38, Last sequence update)  
 16-OCT-2001 (Rel. 40, Last annotation update)  
 Diaphanous protein homolog 1 (Diaphanous-related formin 1) (DRF1)  
 DE (MDIA) (p140mdia).  
 GN DRAPH1 OR DIAP1.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sclerogasthi; Muridae; Murinae; Mus.  
 CC NCBI\_TaxID=10090;  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 MEDLINE=97357293; PubMed=9214622;  
 Watanabe N., Madanle P., Reid T., Ishizaki T., Watanabe G.,  
 Kariizuka A., Saito Y., Nakao K., Jockusch B.M., Narumiya S.;  
 "p140mdia, a mammalian homolog of Drosophila diaphanous, is a target  
 RT protein for Rho small GTPase and is a ligand for profilin."  
 RL EMBO J. 16:3044-3056(1997).  
 RN [2]  
 RP FUNCTION.  
 MEDLINE=20142655; PubMed=10678165;  
 Tomlinaga T., Sahai E., Chardin P., McCormick F., Courtneidge S.A.,  
 Alberts A.S.;  
 "Diaphanous-related formins bridge Rho GTPase and Src tyrosine kinase  
 RT signaling."  
 RL Mol. Cell 5:13-25(2000).  
 CC -1 FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN  
 CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE  
 CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,  
 CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE  
 CC SERUM RESPONSE FACTOR. DER PROTEINS COUPLE RHO AND SRC TYROSINE  
 CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS.  
 CC -1 SUBCELLULAR LOCATION: MEMBRANE RUFILES, ESPECIALLY AT THE TIP OF  
 CC RUPELES, OF MOTILE CELLS.  
 CC -1 TISSUE SPECIFICITY: UBICITOUS.  
 CC -1 DOMAIN: DRES ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE  
 CC RHO-GTP ACTIVATES THE DRES BY DISRUPTING THE GBD-DAD INTERACTION.  
 CC -1 SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).  
 CC -1 SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.  
 CC -1 SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.  
 CC -1 SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.  
 CC -1 SIMILARITY: CONTAINS 1 DRP AUTOREGULATORY DOMAIN (DAD).  
 CC -1 SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. DIAPHANOUS  
 CC SUBFAMILY.  
 CC -----

Query Match	12.1%;	Score 145.5;	DB 1;	Length 1255;
Best Local Similarity	30.0%;	Pred. No. 0.076;		
Matches 66;	Conservative 7;	Mismatches 80;	Indels 67;	Gaps 11

P13608	028159	B79117
<b>STANDARD;</b>		
<b>PGCA BOVIN</b>		
<b>BOVIN</b>		
<b>LT 10</b>		
<b>PRT;</b>	<b>2364</b>	<b>AA</b>

DT\* 01-JAN-1990 (Rel. 13, Created)  
 DT\* 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core  
 protein) (CSPCP).  
 GN AGC1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RN Hering T.M., Kollar J., Huynh T.D.;  
 RA Submitted (Sep-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE OF 563-1056 FROM N.A.  
 RP MEDLINE=89380219; PubMed=5528543;  
 RP Antonsson P., Heinegaard D., Oldberg A.;  
 RA "The keratan sulfate-enriched region of bovine cartilage  
 RT consists of a consecutively repeated hexapeptide motif\*";  
 RL J. Biol. Chem. 264:16170-16173(1989).  
 RN [3]

```

RX SEQUENCE OF 1609-2113 AND 2151-2364 FROM N.A.
RX MEDLINE=87270630; PubMed=31114460;
RA Oldberg A., Antonsson P., Heinegaard D.;
RT "The partial amino acid sequence of bovine cartilage proteoglycan,
RT deduced from a cDNA clone, contains numerous Ser-Gly sequences
RT arranged in homologous repeats.";
RL Biochem. J. 243:255-259(1987).
RN [4]
RP SEQUENCE OF 2114-2150 FROM N.A.
RC TISSUE=cartilage;
RX MEDLINE=93352525; PubMed=8349621;
RA Fuelep C., Walcz E., Valyon M., Gant T.T.;
RT "Expression of alternatively spliced epidermal growth factor-like
RT domains in aggregates of different species. Evidence for a novel
RT module.";
RL J. Biol. Chem. 268:17377-17383(1993).
RN [5]
RP PARTIAL SEQUENCE.
RX MEDLINE=85027710; PubMed=64489519;
RA Perin J.P., Bonnet F., Jolles J., Jolles P.;
RT "Sequence data concerning the protein core of the cartilage
RT proteoglycan monomers. Characterization of a sequence allowing the
RT synthesis of an oligonucleotide probe.";
RL FEBS Lett. 176:37-42(1984).
RN [6]
RP PARTIAL SEQUENCE.
RX MEDLINE=87005253; PubMed=3530809;
RA Perin J.P., Bonnet F., Jolles P.;
RT "Structural relationship between link proteins and proteoglycan
RT monomers.";
RL FEBS Lett. 206:73-77(1986).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGEOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUGCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE C-TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
CC O-LINKED (ABOUT 40) OLIGOSACCHARIDES.
CC -1- PTM: THE KERATAN SULFATE CONTENTS DIFFER CONSIDERABLY BETWEEN
CC ADULT AND FETAL BOVINE PROTEOGLYCAN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U76615; AAB38524.1; -.
CC EMBL: L07053; -. NOT_ANNOTATED_CDS.
CC
CC PIR: A27752; A27752.
CC PIR: A29164; A29164.
CC PIR: B29164; B29164.
CC PIR: E29164; E29164.
CC PIR: G27751; G27751.
CC HSSP: P08709; 1BF9.
CC
CC InterPro: IPR000152; Asx_hydrroxy1.

```



DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR003599; Ig.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR000538; Link.  
 DR InterPro: IPR003324; SGXSG.  
 DR InterPro: IPR000436; Sushi\_SCR\_CCP.  
 DR InterPro: IPR001304; lectin\_c.  
 DR Pfam: PF000059; lectin\_c.1.  
 DR Pfam: PF02339; SGXSG; 62.  
 DR Pfam: PF00084; sushi; 1.  
 DR Pfam: PF00193; xlink; 4.  
 DR ProDom: PD000918; Link; 4.  
 DR SMART: SM00032; CCP; 1.  
 DR SMART: SM00034; CLECT; 1.  
 DR SMART: SM00179; EGF\_CA; 1.  
 DR SMART: SM00409; Ig; 1.  
 DR SMART: SM00445; LINK; 4.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE: PS00022; EGF\_1; 1.  
 DR PROSITE: PS01187; EGF\_CA; 1.  
 DR PROSITE: PS00280; IG\_MHC; FALSE\_NEG.  
 DR PROSITE: PS01241; LINK; 4.  
 DR PROSITE: PS00615; C\_TYPE\_LLECTIN\_1; 1.  
 DR PROSITE: PS50041; C\_TYPE\_LLECTIN\_2; 1.  
 DR GlycoProfile: Cartilage; Proteoglycan; Lectin; Signal; Sushi;  
 KW EGF-like domain; Calcium; Alternative splicing; Repeat;  
 KM Immunoglobulin domain.  
 FT SIGNAL 1 16  
 FT CHAIN 17 2364  
 FT DOMAIN 17 2364  
 FT DOMAIN 170 247  
 FT DOMAIN 268 349  
 FT DOMAIN 504 581  
 FT DOMAIN 602 683  
 FT DOMAIN 774 907  
 FT DOMAIN 1433 2112  
 FT DOMAIN 2113 2149  
 FT DOMAIN 2114 2364  
 FT DOMAIN 2161 2276  
 FT DOMAIN 2280 2338  
 FT DISULFID 51 133  
 FT DISULFID 173 246  
 FT DISULFID 199 220  
 FT DISULFID 273 348  
 FT DISULFID 297 318  
 FT DISULFID 509 580  
 FT DISULFID 533 554  
 FT DISULFID 607 682  
 FT DISULFID 631 652  
 FT DISULFID 1117 2128  
 FT DISULFID 2182 2274  
 FT DISULFID 2250 2266  
 FT DISULFID 2281 2324  
 FT DISULFID 2310 2337  
 FT CARBOHYD 126 126  
 FT CARBOHYD 239 239  
 FT CARBOHYD 333 333  
 FT CARBOHYD 387 387  
 FT CARBOHYD 611 611  
 FT CARBOHYD 667 667  
 FT VARSPLIT 2114 2150  
 SQ SEQUENCE 2364 AA: 246359 MW: 6683763420C3D4C CRC64:

Query Match 12.1%; Score 145.5; DB 1; Length 2364;  
 Best Local Similarity 27.0%; Pred. No. 0.13; Mismatches 75; Gaps 17;  
 Matches 64; Conservative 27; Indels 71; Indels 75; Gaps 17;  
 5 WPOSGLPIHPINKISSTPECTGD---RSVFC---QMEVLDRCYSIPFHRLLC--VSCI 55

DB 635 WLAGSLRPL--VTPRACGGDKPGVRYTVIKNPQGLD---PLSRHNAFCRGVSA 689  
 QY 56 -----KRASGPNPGD-----PG-----PTSLPPS-----TP 78  
 DB 690 PSPEEEKSAPTAGPDVEWVTVQVGPVAVPAGIEETIIPGTVEPENKTEMLAYTP 749  
 QY 79 GSPLPGQDDPADAAEPFGKPTG---SEDHOG-RATOLPGALDTSPPGTOHPFAPEPTIPG 135  
 DB 750 AGTLPLPGIP-----PTWPTGEXTHEHTGTPSATEVPSASEKPP-SEEPFPEEPF 802  
 QY 136 ASWSISPTTGGPLPWGWTOTPTVPEDKQPGEDLRHGTSLPADLGRPPECPHT 192  
 DB 803 ---SEKPPPEPEL-----FPSEKPPSEKPPSEB-----FPSEKPPPEELFPs 845  
 RESULT 11  
 PRP1\_HUMAN STANDARD; PRT; 331 AA.  
 ID PRP1\_HUMAN  
 AC P04280;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Salivary proline-rich protein precursor (Clones CP3, CP4 and CP5)  
 DE [Contains: Basic peptide IB-6; Peptide P-H].  
 GN PRP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=85289325; PubMed=2993301;  
 RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
 RT "Differential RNA splicing and post-translational cleavages in the  
 RT human salivary proline-rich protein gene system.";  
 RT J. Biol. Chem. 260:11123-11130(1985).  
 RL [2]  
 RN SEQUENCE OF 214-331.  
 RP MEDLINE=86243355; PubMed=3521730;  
 RA Kaufman D., Hofmann T., Bennick A., Keller P.;  
 RT "Basic proline-rich proteins from human parotid saliva: complete  
 RT covalent structures of proteins IB-1 and IB-6.";  
 RT Biochemistry 25:2387-2392(1986).  
 RN [3]  
 RP SEQUENCE OF 276-331.  
 RA MEDLINE=84161824; PubMed=6671974;  
 RA Saitoh E., Iseura S., Sanada K.;  
 RT "Further fractionation of basic proline-rich peptides from human  
 RT parotid saliva and complete amino acid sequence of basic proline-rich  
 RT peptide P-H.";  
 RL J. Biochem. 94:1991-1997(1983).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: K03204; AAA60185.1; -  
 CC EMBL: K03205; AAA60186.1; -  
 CC EMBL: K03206; AAA60187.1; -  
 CC PIR: A03291; PIHUB6.  
 CC PIR: C25372; C25372.  
 CC MIM: 180989; -  
 KW Repeat; Parotid gland; Multigene family; Saliva; Signal.  
 FT SIGNAL 1 16  
 FT CHAIN 214 331  
 FT VARSPLIT 276 331  
 FT VARIANT 106 238  
 FT MISSING (IN CLONE CP-4).  
 FT FTID-VAR\_005561.





RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 [16]  
 RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.  
 RX MEDLINE=93293968; PubMed=8514866;  
 RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleinert C.,  
 RA Barley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,  
 RA Cole C.W., Jaakola P., Ryyanen M., Pearce W.H., Yao J.S.T.,  
 RA Majamaa K., Smilens S.V., Gatalica Z., Ferrell R.E., Jimenez S.A.,  
 RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;  
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations  
 RT in the triple-helical domain of type III procollagen are an  
 RT infrequent cause of aortic aneurysms.";  
 RL J. Clin. Invest. 91:2539-2545(1993).  
 [17]  
 RP VARIANT THR-698.  
 RX MEDLINE=91045136; PubMed=2235526;  
 RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,  
 Wu Y., Ganguly A., Prockop D.J.;  
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";  
 RL Nucleic Acids Res. 18:6180-6180(1990).  
 [18]  
 RP VARIANT AORTIC ANEURYSM ARG-786.  
 RX MEDLINE=91056145; PubMed=2243125;  
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;  
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family  
 RT with aortic aneurysms.";  
 RL J. Clin. Invest. 86:1465-1473(1990).  
 [19]  
 RP VARIANT EDS-IV ARG-828.  
 RX MEDLINE=94016385; PubMed=8411057;  
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;  
 RT "The substitution of glycine 661 by arginine in type III collagen  
 RT produces mutant molecules with different thermal stabilities and  
 RT causes Ehlers-Danlos syndrome type IV.";  
 RL J. Med. Genet. 30:690-693(1993).  
 [20]  
 RP VARIANT EDS-IV SER-957.  
 RX MEDLINE=89109135; PubMed=2492273;  
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;  
 RT "A single base mutation that substitutes serine for glycine 790 of  
 RT the alpha 1 (III) chain of type III procollagen exposes an arginine  
 RT and causes Ehlers-Danlos syndrome IV.";  
 RL J. Biol. Chem. 264:1349-1352(1989).  
 [21]  
 RP VARIANT EDS-IV VAL-960.  
 RX MEDLINE=95266429; PubMed=7749417;  
 RA Tromp G., de Paeppe A., Nuytlinck L., Madhathari S.L., Kuivaniemi H.;  
 RT "Substitution of valine for glycine 793 in type III procollagen in  
 RT Ehlers-Danlos syndrome type IV.";  
 RL Hum. Mutat. 5:179-181(1995).  
 [22]  
 RP VARIANT EDS-IV GLU-1014.  
 RX MEDLINE=92316511; PubMed=1352273;  
 RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,  
 Pope F.M.;  
 RT "A single base mutation in the gene for type III collagen (COL3A1)  
 RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos  
 RT syndrome type IV. An unaffected family member is mosaic for the  
 RT mutation.";  
 RL Hum. Genet. 89:414-418(1992).  
 [23]  
 RP VARIANT EDS-IV ASP-1050.  
 RX MEDLINE=90037070; PubMed=2808425;  
 RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;  
 RT "Single base mutation in the type III procollagen gene that converts  
 RT the codon for glycine 883 to aspartate in a mild variant of  
 RT Ehlers-Danlos syndrome IV.";  
 RL J. Biol. Chem. 264:19313-19317(1989).  
 [24]  
 RP VARIANT EDS-IV VAL-1077.  
 RX MEDLINE=91374480; PubMed=1895316;

RA Richards A.J., Lloyd J.C., Ward P.N., de Paeppe A., Narcisi P.,  
 RA Pope F.M.;  
 RT "Characterisation of a glycine to valine substitution at amino acid  
 RT position 910 of the triple helical region of type III collagen in a  
 RT patient with Ehlers-Danlos syndrome type IV.";  
 RL J. Med. Genet. 28:458-463(1991).  
 [25]  
 RP VARIANT EDS-IV GLU-1173.  
 RX MEDLINE=93022543; PubMed=1357232;  
 RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;  
 Query Match 12.0%; Score 145; DB 1; Length 1466;  
 Best Local Similarity 26.1%; Pred. No. 0.092;  
 Matches 72; Conservative 19; Mismatches 71; Indels 114; Gaps 19;  
 QY 2 EGQWVPGSGPLPIPKISTEPCTG-----DRYVF-----COMENVLDNR----- 39  
 DB 7 KQSWL-LIALHPITLLAQEAVEGGCSHLQSYADRVKMRPEQCICVDSGSLCDI 65  
 QY 40 YC-----SIPGYHRLCCVSCIRKASGP-----NPGPDGPTSLP----- 73  
 DB 66 ICDDDELCPNPEIP--FGECCAVCPPTPTPTPRNGQPGQPKGDGPPGIGRNGDP 123  
 QY 74 -----PFSTPGSPDP-----GPD-----PADAEPP 95  
 DB 124 GIPGP-GSPGSPGPGICESCPTGPQNYSPQYDSYDKSGVAVGLAGYGPAGPPGP 182  
 QY 96 GKPTGSEDHQHRATQTLGALDTSRQHP-----FAPETPIPGASMSIPTPGGLP 149  
 DB 183 G-PPTSGH-----PCS--PGSPYQGPGEFGQAGPSP--PGRCGALGSPGAKGD 230  
 QY 150 WGMOTPTPTVPEDKQPGEDLRHPTSLPADLPGR 185  
 DB 231 -GESGRPR-PEGRLPGP---PGIKGPAGIPGP 260  
 RESULT 13  
 ID CAC3\_HAECO STANDARD: PRT; 295 AA.  
 AC P16253;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Cuticle collagen 3A3.  
 GN 3A3.  
 OS Haemonchus contortus (Barber pole worm).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
 OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.  
 OX NCBI\_TaxID=6289;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=90136718; PubMed=2615789;  
 RA Shamsansky L.M., Pratt D., Boisvenue R.J., Cox G.N.;  
 RT "Cuticle collagen genes of Haemonchus contortus and Caenorhabditis  
 RT elegans are highly conserved.";  
 RL Mol. Biochem. Parasitol. 37:73-86(1989).  
 CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 CC MISCELLANEOUS: THIS PROTEIN SHOWS A POTENTIAL TRIPLE-HELICAL  
 CC REGIONS, WHICH CONTAIN GLYCINE AS EVERY THIRD AMINO ACID.  
 CC -1- MISCELLANEOUS: IN ALL NEMATODE CUTICLE COLLAGENS, THE POLYPEPTIDE  
 CC CHAINS ARE COMPLEXED WITHIN THE CUTICLE BY DISULFIDE BONDS AND  
 CC OTHER TYPES OF COVALENT CROSS-LINKS.  
 CC -1- SIMILARITY: TO CAENORHABDITIS ELEGANS CUTICLE COLLAGENS.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----  
 DR EMBL: M32821; AAA29174.1; -  
 DR EMBL: M32820; AAA29173.1; -  
 DR InterPro: IPR002486; Col\_cuticle\_N.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 2.  
 DR Pfam: PF01484; Col\_cuticle\_N; 1.  
 DR Cuticle; Connective tissue; Repeat; Multigene family; Collagen.  
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.  
 SQ SEQUENCE 295 AA; 29618 MW; 85355A431B6D331 CRC64;

Query Match 11.8%; Score 142; DB 1; Length 295;  
 Best Local Similarity 34.3%; Pred. No. 0.037;  
 Matches 58; Conservative 7; Mismatches 50; Indels 54; Gaps 15;

51 CVCICKASGPPGPPGPTSLP--PFTPGSP-LPG-----PQD-----P 88  
 91 CESCGR--PGP-PGP-PGPPGPPGP-GRPGAGLPVGLPPPGSCPPVSIIPCAACP 145  
 89 ADAEPPGKP---TGSEDHQGRATQ--LPGALDTSPTQHPAPETP-LPGASWST- 140  
 146 AGPPPGPKGPPGPPGDEGQPPGQDGLPGQGTGKP---PGPPGPKGPPGSGEYG 201  
 QY 141 -----SPTTPGGLPKGWTQTPTPVPRDKQPPEDLRHNETSLRADLPGR 184  
 202 EDAECPPVAPG-----DQGPPEPPGPPGPPG---PGLQGPVGMPCQ 240  
 Db

RESULT 14  
 CC12-CAEEL STANDARD; PRT; 316 AA.  
 AC P20630;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Cuticle collagen 12 precursor.  
 GN COL-12 OR F15H10.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=90172409; PubMed=1689778;  
 \*Tandemly duplicated Caenorhabditis elegans collagen genes differ in  
 their modes of splicing.\*  
 J. Mol. Biol. 211:395-406(1990).

[2]  
 SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Berks M.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 CC LINKS.  
 CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE  
 CC COLLAGENS. THE MATURE COL-12 AND COL-13 ARE IDENTICAL.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X51622; CAA35954.1; -  
 DR EMBL: Z73972; CAA98257.1; -

DR PIR: S08169; S08169.  
 DR Wormpep; F15H10.1; CE05638.  
 DR InterPro: IPR002486; Col\_cuticle\_N.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 2.  
 DR Pfam: PF01484; Col\_cuticle\_N; 1.  
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen;  
 FT Signal.  
 FT SIGNAL 1 36  
 FT CHAIN 37 316  
 FT DOMAIN 128 157 CUTICLE COLLAGEN 12.  
 FT DOMAIN 176 202 TRIPLE-HELICAL REGION.  
 FT DOMAIN 206 235 TRIPLE-HELICAL REGION.  
 FT DOMAIN 240 266 TRIPLE-HELICAL REGION.  
 FT DOMAIN 269 304 TRIPLE-HELICAL REGION.  
 SQ SEQUENCE 316 AA; 30098 MW; 6CA81FF94706D42E CRC64;

Query Match 11.7%; Score 141.5; DB 1; Length 316;  
 Best Local Similarity 27.8%; Pred. No. 0.041;  
 Matches 58; Conservative 13; Mismatches 55; Indels 83; Gaps 13;

51 CVCICKASGPPGPPGPTSLP--PFTPGSP-LPG-----PQD-----P 70  
 122 CCGSGGAGPPGPPGPPGQDAPGNDAPGAPGPPGQDASPDQAGDSFCFCPPAGPPPS 181  
 71 SLP-----PFTPGSP-----LPGPPDADAEPPGKPTGSEDH-----QHGATOLPG 114  
 182 GAPGQKPGSAPGAPGAPGQSGALPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPG 240  
 QY 115 ALDTSSPGTQHPAPETPLPGASWSTISPTTPGGLPWGWTQTPTPPE-DKQPPEDLRHP 173  
 241 -----TPGAPG--PESPPAG--APQPPQA--GSSQPPGPPGQDADAPGA---P 283  
 QY 174 GTSPLADLPGRPP--CHPTGT 194  
 284 GAPGQAGAPGQDGGSSGACDHCPRPT 312  
 Db

RESULT 15  
 CC13-CAEEL STANDARD; PRT; 316 AA.  
 AC P20631;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Cuticle collagen 13 precursor.  
 GN COL-13 OR F15H10.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=90172409; PubMed=1689778;  
 RA Park Y.-S.; Kramer J.M.;  
 RT \*Tandemly duplicated Caenorhabditis elegans collagen genes differ in  
 RT their modes of splicing.\*  
 J. Mol. Biol. 211:395-406(1990).

[2]  
 SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Berks M.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 CC LINKS.  
 CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE  
 CC COLLAGENS. THE MATURE COL-12 AND COL-13 ARE IDENTICAL.

DR EMBL: X51622; CAA35954.1; -  
 DR EMBL: Z73972; CAA98257.1; -





FT FTID=VAR\_008152.  
 FT CUTICLE 1030 1030 G-> V (IN AS).  
 FT VARIANT 1201 1201 /FTID=VAR\_008153.  
 FT VARIANT 1402 1402 G-> S (IN AS).  
 FT VARIANT 1572 1572 /FTID=VAR\_008154.  
 FT VARIANT 1659 1659 /FTID=VAR\_008155.  
 FT CONFLICT 1690 AA; 164095 MW; E1E72F283A72BAAE CRC64;  
 SO SEQUENCE

Query Match 11.7%; Score 140.5; DB 1; Length 1690;  
 Best Local Similarity 35.7%; Pred. No. 0.19;  
 Matches 55; Conservative 2; Mismatches 56; Indels 41; Gaps 11;

QY 58 ASGPNG-----PDGPTSLPFGSTPGSP-LPGPDADAAEPKPTGSEDH 104  
 DB 1237 SSGP-PGPGATGAKRAKDIIPDEP-----PGDGPFGPGPGKAPGPGGLP-GSVDL 1286  
 105 QHGRATQLPGALDTSSPGTQHPPEP-----PGASWSISPTTPTG--PWGWTQPT 157  
 1287 LRGE-----PG-DCGLPGPPGPGPGPGKPGCGDKGKPGWPGPGQPGHGPFG 1340  
 QY 158 PVEDKGPGEEDLRHPTGLPADLGRPRPECHP 191  
 DB 1341 P-PGKGLPGP-----PGRKGPGLPGRPGSPPP 1369

RESULT 18  
 CC34\_CAEEL STANDARD; PRT; 298 AA.

AC P34687;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE Cuticle collagen 34.  
 GN COL-34.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=93013043; PubMed=1398138;  
 RA Bird D.M.;  
 RT "Sequence comparison of the Caenorhabditis elegans dpy-13 and col-34  
 genes, and their deduced collagen products.";  
 RL Gene 120:261-266(1992).  
 -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 LINKS.  
 -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE  
 COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M80650; AAA27985.1; -  
 CC PIR: JCI448; JCI448.  
 CC InterPro: IPR002486; COL\_cuticle\_N.  
 CC InterPro: IPR000087; Collagen.  
 CC Pfam: PF01391; Collagen; 2.

DR Pfam: PF01484; COL\_cuticle\_N.1.  
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.  
 FT DOMAIN 103 132 TRIPLE-HELICAL REGION.  
 FT DOMAIN 151 177 TRIPLE-HELICAL REGION.  
 FT DOMAIN 181 198 TRIPLE-HELICAL REGION.  
 FT DOMAIN 215 277 TRIPLE-HELICAL REGION.  
 SO SEQUENCE 298 AA; 29653 MW; 87E5370E0172D182 CRC64;

Query Match 11.5%; Score 139; DB 1; Length 298;  
 Best Local Similarity 38.4%; Pred. No. 0.055; 58; Indels 22; Gaps 11;  
 Matches 53; Conservative 5; Mismatches 58;

QY 60 GNPDPGPTSLPFGSTPGSP-----LPGPDADAAEPKPTGSEDHQHRATQ 111  
 DB 151 GP-PGP-PGPGPGDSDGSGPSGLPGQDAAPGPGKPGKPGAP-GAPEHQSECEP 207  
 QY 112 LGALDTSSPGTQHPPEP-----PGASWSISPTTPTG--GLPWGWTQPTTPVEDKGPGEED 169  
 DB 208 RGEPLRPGEPGPGAGPGPG-PG-----SPGPGADGSP-GDPGKPGNGPD-GPGAD 259  
 QY 170 LRHPTGLPADLGRPRPE 187  
 DB 260 -GNPAGPAGPPGSPGE 276

RESULT 19  
 CC34\_CAEEL STANDARD; PRT; 302 AA.

AC P17657;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Cuticle collagen dpy-13.  
 GN DPY-13 OR DPY-16 OR F30B5.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX WATERSTON R.;  
 RA Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RL  
 RN  
 RN VARIANT DPY13(E225).  
 RP MEDLINE=93013043; PubMed=1398138;  
 RA Bird D.M.;  
 RT "Sequence comparison of the Caenorhabditis elegans dpy-13 and col-34  
 genes, and their deduced collagen products.";  
 RL Gene 120:261-266(1992).  
 -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 LINKS.  
 -1- DISEASE: MUTATIONS IN DPY-13 AFFECTS THE BODY SHAPE.  
 -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE  
 COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@sib.ch](mailto:license@sib.ch))

```

CC      or send an email to license@sb-sib.ch).
CC      -----
DR      EMBL; M23559; AAA27994.1; -.
DR      EMBL; U42437; AAB83499.1; -.
DR      PIR; A31921; A31921.
DR      Wormpep; F30B5.1; CE00462.
DR      InterPro; IPR002486; Col_cuticle_N.
DR      InterPro; IPR000087; Collagen.
DR      Pfam; PF01391; Collagen; 2.
DR      Pfam; PF01484; Col_cuticle_N; 1.
KT      Cuticle; Connective tissue; Repeat; Multigene family; Collagen
FT      DOMAIN      106      135      TRIPLE-HELICAL REGION.
FT      DOMAIN      154      210      TRIPLE-HELICAL REGION.
FT      DOMAIN      219      278      TRIPLE-HELICAL REGION.
FT      VARIANT      205      278      G -> E (IN DPLY34E225))
FT      SEQUENCE     302 AA;  30052 MW;  16B5CDAP14E0F0E5 CRC64;

```

Query Match	11.58;	Score 138.5;	DB 1;	Length 302;
Best Local Similarity	29.68;	Pred. No. 0.059;		
Matches 58; Conservative	7;	Mismatches 62;	Indels 69;	Gaps 12;

[illegible]

RESULT 20	
DRPL_RAT	
ID	STANDARD;
DRPL_RAT	PRT; 1183 AA.

AC P54258;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DT Atrophin-1 (Dentatorubral-pallidumlysin atrophy protein).  
DRPLA.  
NCBI TaxId=10116;  
OC Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RX NCBI\_TaxId=10116;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Cerebellum, and Striatum;  
RC MEDLINE=96081227; PubMed=9173996;  
RX Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,  
RA Ashworth R.G., Ross C.A.;  
RT "Cloning and expression of the rat atrophin-I (DRPLA disease gene)  
RT homologue".  
RT Neurobiol. Dis. 2:129-138(1995).  
RN [12]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;  
RC MEDLINE=96081227; PubMed=8541849;  
RA Schmitt I., Epplen J.T., Riese O.;  
RT "Predominant neuronal expression of the gene responsible for  
RT dentatorubral-pallidumlysin atrophy (DRPLA) in rat".;  
RT Hum. Mol. Genet. 4:1619-1624(1995).  
CC -1- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH  
CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.  
CC -1- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES  
CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).  
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL; U31777; AAA80337.1; -.
DR	EMBL; X89453; CAA61623.1; -.
DR	InterPro; IPR002951; Atrophin.

[illegible]

Query Match	11.5%;	Score 138.5;	DB 1;	Length 1183;
Best Local Similarity	36.7%;	Needle	Yes	0.10

Best Local Similarity 26.7%; Pred. No. 0.18;  
Matches 55; Conservative 18; Mismatches

6 YPOSGPIHPNKI\$STEBCTCDBSEVEGCOMETI DBVCSIDCV - - - - -HBT COWCSTIWAACCN 61

QY	62	NCPDGGPGLSPLEPTPCSLRGPQPDADAEPFGK-PRGSEHNGRATQLPGALDTSS	122
Dd	302	LPPPALRLNNASAPRGMAQ-----PIGHILPSRHAMG-----QGSGGL	343
Dd	344	PPEPEGKPTLAP-----SPIHLPPASSAPAGPMRYPVSSCSSSVAASSSSSATSQY	397
QY	121	PGNQ-----HEAEPTITCASMSIS-----PLTGGGLPWMTQTTPPYVEDKGQR	166
Dd	398	PASQILPSTYPHSFEPPT-----SMSVSNQPKYTOSTLSQAWSQAGPPPPPPGRLLP	452
QY	167	GEDLRHPTGLSP--ADLRGRPRECH	190
Dd	453	NNN--THPGPPPTGGOSTAHPPAFAH	477

Search completed: August 29, 2002, 00:12:01  
Job time: 332 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 20:09:28 ; Search time 87.84 Seconds

(without alignments)  
2465.729 Million cell updates/sec

Title: US-09-938-330-20

Perfect score: 6954

Sequence: 1 MAPLRALLSYLLPLHCALCT.....PTGFTLVCVLRDSQLRQHT 12552

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP virus:\*  
16: SP bacteriophage:\*  
17: SP archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1072.5	15.4	2165	5	Q19791
2	1038.5	14.9	1054	5	Q9W493
3	962	13.8	1229	5	Q9V61
4	948	13.6	1427	4	Q96L37
5	810	11.6	364	4	Q96AV5
6	774.5	11.1	550	4	Q9NSJ8
7	707	10.2	2174	5	Q9G0R0
8	707	10.2	3060	5	Q9YAV4
9	699.5	10.1	2167	5	Q76840
10	683.5	9.8	1235	4	Q95428
11	681.5	9.8	1280	11	Q9EPX2
12	677	9.7	1444	5	Q17591
13	668.5	9.6	3198	5	Q9U8G8
14	663.5	9.5	1572	5	Q44938
15	622	8.9	951	4	Q60345
16	614	8.8	1062	5	Q19204

17	597.5	8.6	525	4	Q96RW4	Q96RW4 homo sapien
18	595.5	8.6	766	4	P82987	P82987 homo sapien
19	556.5	8.0	860	5	Q22580	Q22580 caenorhabdi
20	518	7.4	1014	5	Q95R33	Q95R33 drosophila
21	515	7.4	1091	5	Q9W126	Q9W126 drosophila
22	480.5	6.9	269	6	Q9GL55	Q9GL55 drosophila
23	433	6.2	898	4	Q9FE24	Q9FE24 homo sapien
24	430	6.2	1059	5	P90884	P90884 caenorhabdi
25	427	6.1	364	4	Q9UGQ1	Q9UGQ1 homo sapien
26	388.5	5.6	1637	6	Q9XSX8	Q9XSX8 bos taurus
27	384.5	5.5	610	13	Q93523	Q93523 bothrops ja
28	379	5.5	227	4	Q9HX28	Q9HX28 homo sapien
29	379	5.5	340	11	Q91256	Q91256 mus musculu
30	362.5	5.2	610	13	Q9120	Q9120 agkistrodon
31	361	5.2	609	13	Q90282	Q90282 crocalus at
32	353	5.1	728	4	Q9B211	Q9B211 homo sapien
33	350.5	5.0	873	13	Q42595	Q42595 xenopus lae
34	342.5	4.9	918	4	Q9H013	Q9H013 homo sapien
35	342	4.9	918	4	Q9B215	Q9B215 homo sapien
36	341	4.9	920	11	Q35674	Q35674 mus musculu
37	339.5	4.9	616	13	Q90495	Q90495 echis carin
38	337	4.8	609	13	Q9W6M5	Q9W6M5 agkistrodon
39	332	4.8	1023	5	Q967H9	Q967H9 strongyloce
40	325	4.7	857	4	Q9UHP2	Q9UHP2 homo sapien
41	323.5	4.7	4123	4	Q75851	Q75851 homo sapien
42	320.5	4.6	1569	5	Q9VFR0	Q9VFR0 drosophila
43	319	4.6	1081	5	Q90631	Q90631 drosophila
44	318.5	4.6	5636	4	Q96RW7	Q96RW7 homo sapien
45	317.5	4.6	622	4	Q9H8X0	Q9H8X0 homo sapien

## ALIGNMENTS

RESULT 1  
ID Q19791 PRELIMINARY: PRT: 2165 AA.  
AC 019791: 027524:  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE F25H8.3 PROTEIN.  
GN F25H8.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;  
OC Rhabditiida; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gajdarsky S.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Alnsough R., Anderson K., Haynes C., Berts M.,  
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Riffen L., Roopra A., Saunders D., Showkeen R.,  
RA Smither-Mieg J., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
ellegans.";  
RL Nature 368:32-38(1994).  
DR EMBL: Z69361: CA93288.1;  
DR EMBL: Z69360: CA93288.1;  
DR EMBL: Z69360: CA93287.1;  
DR EMBL: Z69360: CA93287.1;  
DR HSSP: P15167: IDTH.  
DR MEROPS: M12.135: -.

DR InterPro: IPR001590; Reprolysin.  
 DR InterPro: IPR000884; TSP1.  
 DR Pfam: IPR000130; Zn\_MTPetase.  
 DR Pfam: PF01421; Reprolysin. 1.  
 DR Pfam: PF00090; TSP1; 14.  
 DR SMART: SM00209; TSP1; 18.  
 DR PROSITE: PS50215; ADAM\_MERO. 1.  
 DR PROSITE: PS50092; TSP1; 6.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN. 1.  
 DR SEQUENCE 2165 AA; 244397 MW; FCC3DAB8AAA9C4888 CRC64;  
 Query Match 15.4%; Score 1072.5; DB 5; Length 2165;  
 Best Local Similarity 26.0%; Pred. No. 7,4e-75;  
 Matches 312; Conservative 176; Mismatches 425; Indels 287; Gaps 45;  
 151 ECYVTGGVTPGMPGAVALISNC---DGLAGLIRDTSDFTIEPLERG---OOEKEASGRTH 204  
 168 DCITRAHVKGVHONST-VNLCDSFDGLXGMLALPSPGITHYEPITSGNGTEHDGASRRH 226  
 205 VYTRREAVQOEWAEPDGDLDNEAFGLDLPNLGLGVDDQGDGT-ERK---RRHAKPGSYS 260  
 227 LVRFEDPMEHFEKSPD-----HINSTSVNETETTVATWODOWEDYIERKARRRANSWDHY 281  
 261 IEVLIVYDDSVYRPHGKEHYONVYLTLMNIVDEIYHDESGVHNITLAVRLIMVGYRSL 320  
 282 VEVLVADVTMYEYHNS-LEDYVILFSTYASITRHQSLASINAVYVYKILVTKTENAG 340  
 321 SLIERGNFSRLSEQYCRWASHQORODPSHAHNDHVVFLTRDQFSPS-----GYAVPT 373  
 341 PRITQ-NAQOTLDFCFRMOQYNDPDDSSVQHHDVALITLERKIDICRSGKQKDTGLAELG 399  
 374 GCMHPLRSCALNHEDGSSAFVIAHETGHVILGMEHDGQGNCA-----D 417  
 400 TMCMDOKSCALTEEDNGLSAETIAHGLHVSIPHDE-RKCSYPMYKNVCKFOSTKFD 458  
 418 ETSLSGS---WAPLVQAFHFRHNSRCSKLELSRYLPS---YUCLLDPDFDPAWPOP-- 468  
 459 KTOFQNNHIAPILEYTHWMSWSPCSAGMLEFLENRRQTCGLDQPERERYEDVF 518  
 469 -PELPGINYSMEDECRFDESGGYOTCLAFTEPECKOLWC-SHPDNPFCKTKKGPPLDG 526  
 519 VRDEGKKYDHAQOCKEFYGPASELCIPMPY---CRRLMCATFPGSQGCTQTQMPADG 575  
 527 TECAAGK--WCFKGHCIMKSPGQYTGOGGSSWTKGSGSCRSRGGVGRSRSRCSNPS 584  
 576 TPDCSRSMFCHHGACVLAPESLTKIDGQMDKMSMGECGRTCGGGVQKLRCDSPK 635  
 585 AYGRPCLGPMFEYQVNCSECPGTIEDFRAQOACAKRNSYVH---ONAKHSWVP-YEP 639  
 636 RNGKYYGVGQERERYRSCNTGCPMDTQPYREVQCSFENNKDGIQGVASTWTHWPKYAN 695  
 640 DDDAOKCELIIQASDTGVVFMNVOYHGTGRCSRDYPSVARECEVYVGDCKREVSMA 699  
 696 VAPNRCKLYCRLSGSAFYLLRQKVVDGTPCD-RNEDDITVAACMPAGCDHQLHSLR 754  
 700 DDKCGGSDNSHCTVGTGLKASKQAGALVOIPAGAHIOLEAL---ESPSHRSV 755  
 755 RDKCGVCGDDSSCKVAVGTENEDG-TGYNEVAKIPRGSANIDIRKGYNNKMEEDNYL 813  
 756 VKNQVTSFIIINPKGKATSR---TFTAMGLEW-EDAVEDAKSLKTSGLPLPAIALA 810  
 814 SLRANGSEFLNHHQVSLAQAFODTVLEYSGDAI---IRINOTGIRSDIYVH 870  
 811 LPPTGGRRSLAYKY---VIHEDLPLIG-----SNVILE 844  
 871 L--SVGSHPPDISYEYMTAAVNPVIRPISALYLMRTVDTWTECDRACRQOQSKLMCL 928  
 845 EMDTY-----EWALKSWAPCSACAGGAGGIGTOFTYVGGRR 877  
 929 DKSTIRQSHDRNCQNVLPKQATRCMNTIDCSTRMTTEDVSSCSAKCGSGQKQROVSCVKM 988  
 878 RDHNM-VQRIHLCDHKKRPPI----- 897

DB 989 EEDROTPASEHLCDRNSRPSDIASCIYDSSGKKNYGEWTSCESETGSGNKGMRKSYCVD 1048  
 989 -----RRRCNQHPCSQPYWTEWAGCSRSCGKIGVOTRGIQCLLP 938  
 1049 DSNRRVDESLCGRQKEATERECNRIPC--PRWYGHWSRCSGCDG-GVYMRHQAOCIDA 1105  
 939 LSNQTH--KVMAK-----AC-----AGD-----RPE 958  
 1106 ADRETHTSKGAQOEHCHENHACTWMOGWSDCSAKCGDGVYRDANCTDRHRSLYPE 1165  
 959 AR-----RPPCLRVPCPAQMRGLANSOCSATCGEGTIOQROYVCRTNAN-----SLGH 1004  
 1166 HNCIKKEKTIITKPCRHRESP-KYKLGEMSGQCSVEDGMSRRVACVGNGETVDMSLCG 1224  
 1005 CEGDRPDVQVCSL-----PACGQHNSYRADVWEIGTEPGQWVPPSG 1049  
 1225 TASDRPASHQTNLTGCPWRMTDWSACSVCSCGIGHRETEC-IYRQSVASFCGDT- 1282  
 1050 PLHPIKISMSMAEPCTGDR-----SVFCQMEVLDRCSTIPGYHRLCCVSCIKRKSAG 1102  
 1283 ---KMFETSOTCHLLPCTISWKPSSHMSPCSVTGGSGIOTR-----SVSCTRSGSEG 1328  
 RESULT 2  
 ID 09W493 PRELIMINARY; PRT: 1054 AA.  
 AC 09W493;  
 DT 01-MAY-2000 (TRENBLREL. 13, Created)  
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
 DE CG4096 PROTEIN.  
 GN CG4096.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OX Ephydriidae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blaise R.G., Chame M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aguirre A., An H.-J., Andrews-Frankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu U., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya P., Brothier P.,  
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington R.D.C., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,





Matches 284; Conservative 145; Mismatches 400; Indels 172; Gaps 44;

```

QY 249 RRRRAKAGSYIEVLVYDDSVYRFHKEHVNVTITLANYDEIYHDSLSGHINIAL 308
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70  QORRAAGGIIHLELLVAVGDPVQAH-QDETERXYLTINIGELLRDLPSLGAOFVHL 128
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 309 VRLIHWGYROSILIERGNFSPSLQYCRMAHSQRODPFHAHHHVVLTFTQDNG-PS 367
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 VKMVLTFEPGAPNI-TANLTSLSLSCGMSQITNPEDDTDPGHADIVLITTFDELPL 187
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 368 GYAPYTG-----CHPLRSCALNHEDGFSSAFVIAHETGHVLMEDHG-QNGCADETS 420
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 421 LGSVAPLYQAIFHFRHMSRCKLSRLYPS--YDCLLDPPDPAMPQ-----PPEL- 471
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 246 -GHVWASGAAIPRGLAMSPCSRQHLSTLSAGARCV---WDPFRPGSGAHPDPAQ 300
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 472 PGINYSMEQCRFDGSGSYQTC-LAFRTFEPCKQIMC-SHPDNPYCKTKKGPPLDGTG 529
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 301 PGLYASANEQCRVAFGPKAVACTFARHEDLMQALSCHTPRLDQSSCLLVLIDGTGTC 360
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 530 APGWKCFRGHC---IMKSPEDTIGDGGSSWTKFGSCSRSGGCVRSRSCNNPSPAY 586
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 361 GVEKWCSSGRCRSLVELTP--IAAVHGRMSSMGPRSPCSRSGCGGVVTRRRQCNPRPAF 418
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 587 GGRPCLGPMFEYOVCNSEECGTYEDFRAOQCAKN-----SYVHONAKHSWY 635
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 419 GGRACVADLQAEKNTQACEKTOLEMSQOAKRTDQPLRSSPGASAFY-HMGAA--V 474
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 636 PYEPDDAOKCELLQASADTGDVVFNMVYVHDTGRCSTYRDP-----YSVARECEYVPGC 690
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 475 PHSQGD--ALCRHMCRAIGSEFIMKRDSFLDGTROMPSGPRDGLSLICVSSGCTTFC 532
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 691 DKEYGSKADKDCGVCGDSDHCHRTYKG--TLKASKQAGALKVYDIPAGARHIQIEALE 748
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 533 DGRMSQVWDRCOVCGDSDSTSPRGSFTAGAREYVTLFVTP-----NLTSYV 584
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 749 KSPHRSYVKN---QVTSFILNPKGEATSRPTAM-----GLEWEDAVEADAK---ESLK 797
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 585 IANHRPLFTLAVRIGRYVYAGKMSISPTTYPSLLEGGRYKVALIEDRLPRLEETR 644
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 798 TSGPLP-ALAILALPTEGCG--PRSSLAHYKVIHEDLLPLGNNVLEEMDTYEWAL 853
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 645 IWGFLQEDADIOYRRYGEYGNLTPDITFTY-----FQKPRQAWY- 687
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 854 KSWA-----PCSKACGGGIOFTXYGCRRRRDHMHVORHLCDHKRPRIRRCNQHPSCQ 909
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 688 --MAAVGPGSVSCAGLIRWVNSCLDQAKKELVETVQCGSOQPPAMPDEACVLEPC-PP 744
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 910 VVWTEWAGCSRSCKLGVOTRGIOCLPLPSNGTHKVPKACAGDRPEARRP-----C 963
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 745 YMAVGDGPGCSASCGG-GLREPRVCVEAOGSLTKTLTPARACRAG---AQPPAVALLEIC 799
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 964 LRVPCPAOMLIGAMSCSACGEGIOOROVYCTNANSIGHEGDRPDYV-----QVC 1016
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 800 NPQPCPARWVSPSSCTSGAGGLALENETCVPGADGLEAVITGPGSVDEKLPAPEPC 859
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1017 SLPAC--GGNHQNSTVRADWELGTPEQO-----WVPSQPLH----- 1052
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 860 VGMSCPGMCHLDATSGE--KAPSWGSIIRGAQAAHWTVWASCSVSCSGRIGMELRF 917
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1053 -----PINKISSMCAAEPTGDRSVFCOMEVLDYRCSIPGHRIL--CCVSC---I 1097
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 918 LCMDSALRYVDE--ELCGIASKPGSRREVQAVP---CPARWQYKLAACSVSCGRGVV 971
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1098 KKA--GQPNPGDPPTSLPPTSTGSPPLPGHODPADAAEP 1136
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 972 KRLIYKARAHGEDDEEILLDTQCGGLPRPEQD-ACSLER 1011
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT  
096AYS

```

ID 096AYS PRELIMINARY; PRT; 364 AA.
AC 096AYS;
DT 01-DEC-2001 (TREMBLrel, 19, Created)
DT 01-DEC-2001 (TREMBLrel, 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel, 19, Last annotation update)
DE HYPOTHETICAL 40.2 KDA PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA (MEMO CELL LINE);
RA Blum H., Bauersachs S., Mewes H.W., Well B., Wiemann S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162080; CAB82413.1;
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR001128; Cyt_P450.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF0090; tsp_1; 3.

Query Match 11.6%; Score 810; DB 4; Length 364;
Best Local Similarity 43.6%; Pred. No. 2,4e-55;
Matches 157; Conservative 43; Mismatches 76; Indels 84; Gaps 9;

QY 844 EEMDTYEMALKSNAPCSKACGGGIOFTXYGCRRRRDHMHVORHLCDHKRPRIRRCNQ 903
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1 HEDTFEMALKSNAPCSKACGGGFOYTKYGRKRSKNKWHRSFCANKPKPIRRMCNI 60
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 904 HPCSPVWTEWAGCSRSCKLGVOTRGIOCLPLPSNGTHKVPKACAGDRPEARRP 963
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 QECHPIPLWAEWEHCKTKTCGSSGYOLRTYKLOPLDGTNHSVHKTOMGPRPSRRPC 120
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 964 LRVPCPAOMLIGAMSCSACGEGIOOROVYCTNANSIGHEGDRPDYVQVCSLPACGG 1023
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 NRVPCPAOMLIGAMSCSACGEGIOOROVYCTNANSIGHEGDRPDYVQVCSLPACGG 176
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1024 NHQNSTVRADWELGTPEQOVSQPLHPIKISSMCAAEPTGDRSVFCOMEVLDYR 1083
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 -----EPCLDSTFCOMEVLDYR 196
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1084 SIPTGHRILCCVSCIKRAG-----PNPGRPG-----PTSLPFP--STP 1120
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 197 SIPTGHRILCCVSCIKRAG-----PNPGRPG-----PTSLPFP--STP 256
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1121 G-----SPLGPDPPADAAEPCKPTGSEDDHGRATOLPGALDT-----SSPQT 1166
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 AKKMSLSISSVGGPNAYA-AFRPNKSPDGANLRQ--RSAQOAGSKTVRLVTPSSPTK 313
Db      : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
QYNSJ8 PRELIMINARY; PRT; 550 AA.
AC QYNSJ8;
DT 01-OCT-2000 (TREMBLrel, 15, Created)
DT 01-OCT-2000 (TREMBLrel, 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel, 17, Last annotation update)
DE HYPOTHETICAL 60.3 KDA PROTEIN (FRAGMENT).
GN DKEZP762C1110.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA (MEMO CELL LINE);
RA Blum H., Bauersachs S., Mewes H.W., Well B., Wiemann S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162080; CAB82413.1;
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR001128; Cyt_P450.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF0090; tsp_1; 3.

```

DR SMART: SM00209; TSP1: 3.  
 DR PROSITE; PS00339; AA: TRNA\_LIGASE\_II.2; UNKNOWN\_1.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; UNKNOWN\_1.  
 DR PROSITE; PS50092; TSP1: 2.  
 DR Hypothetical protein.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 550 AA; 60337 MW; 97BF7332A54B880 CRC64;

Query Match 11.18; Score 774.5; DB 4; Length 550;  
 Best Local Similarity 30.4%; Pred. No. 2,7e-52;  
 Matches 196; Conservative 89; Mismatches 225; Indels 135; Gaps 28;

QY 413 NGCADENSLGCVAPLVAQAFHFRHRSKCSKLELSRLPS--XDCLLDDEPDAMPQPE 470  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 4 NGVNDSDHM--MASMSNDHSQPSPCSAHYMTTSLDNGHGECLDKQNP1-QLPGD 59  
 QY 471 LRGINYSMEDECRPFDSGYOTCLAFRTFEPCKOLMCSHPDNPYF-CKTKGPRDLGTEC 529  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 60 LPTGSIYDANRQCFTEFGDSKHP--DAATSCSLMCTGTSGGLVYQTKHPMADGTSC 117  
 QY 530 AFGKWFCEGKCIWKSPTDTYQ--DGMSSWTFKFGSCSGGSGGVRNRSRSCNNPSPAYG 587  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 118 GEKWKCIWGCNVNKTDRKHEDTPFHSGMGMGMGWDGSRITCGGGVQYTMRECDNPYKNG 177  
 QY 588 GRPLGPFHFEYQVQNSSECP-GTYEDPRAQCAKRNHYHQAQKHS-----WVY-Y 637  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 178 GKYCEGKRVYRSCNLEDCPDNNGKTRFEQCEAHNE-----SKAFSGSPAVEMLPKY 232  
 QY 638 EPDDDAKCELICOSADTGDVYFMNOVYHDGTRCSYRDPYSCARGCVPYGCDEKVEGSM 697  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 233 AGVSPKRCXKILCOAKIGYEFVLQPKVVDGTPCS-PDSTSVYQGGCVAKAGCDRIIDSK 291  
 QY 698 KADKCGVCGGDNHSCRTVYKTLGKASKQAGALKVQIPAGARIQIQALESKSHRS---754  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 292 KKFDCGCVGGNGSTCKKISGSV--TSAKPGYHDIITPTGATIEVKQNRQSRNNGS 349  
 QY 755 -VVKNOVVSFLNPKGKATSRFTFAMLEWE-----DAVEDAKSLKTSPLP 803  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 350 FLATKADGYIILN-----GDYITSLTEQDITMKGVLYRSGSSALELRIISFSLK 401  
 QY 804 EALIALPTEGEP-RSLAYKYVYIHBDLLPLIGSNVNLLEMDTYEWALKSNAPCSKA 862  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 402 EPLTIQVL--TYGNALRPIKITYFY-----425  
 QY 863 CGGGIQTFTKCGRRRHHMQRHLCHKKRPRPIRRCONHPGSOPIVWTEEGACASRS 922  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 426 -----KKR-----KESFNAP-TFSAWVLEWEGCSKS 452  
 QY 923 CGKLGVTGRIQCLPLSLNGTHKVPAPAKACAG--RPEARRCLRVPCPAQWRLGAMSGCS 981  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 453 C-ELDMQRRLVLC--RDING--QPASECALVKVPASTRCADHPCC-QMQLGEMSSCS 504  
 QY 962 ATGEGIQOROVVCHT-NANSLGHCED---RP-DTYOVCSLPLAC 1021  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 505 KTCGKGYKKRSIKLSHDGVLSHESCDPLKRPKHFIDFCTMARC 549

RESULT 7  
 Q9G0R0 PRELIMINARY; PRT; 2174 AA.

AC Q9G0R0;  
 DT 01-MAR-2001 (Tremblrel. 16, created)  
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)  
 DE EXTRACELLULAR MATRIX PROTEIN PAPILIN PROCURSOR.  
 GN PPN.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=DP CN BW;  
 RX MEDLINE=20530499; PubMed=11076767;  
 RA Kramarov I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,  
 RA Kramarov A.A., Kasche-Gullberg M., Kramer J.M., Ackley B.D.,  
 RA Sieron A.L., Prokop D.J., Fessler J.H.,  
 RT "Papilin in development: a pericellular protein with a homology to the  
 RT ADAMTS metalloproteinases."  
 RL Development 127:5475-5485(2000).  
 DR EMBL: AF205357; AAG37995.1; -  
 DR HSP, P12111; 2KWT.  
 DR Flybase: FBgn000137; Ppn.  
 DR InterPro: IPR003599; 19.  
 DR InterPro: IPR003598; 19\_c2.  
 DR InterPro: IPR003600; 19-like.  
 DR InterPro: IPR002223; Kunitz\_BPTI.  
 DR InterPro: IPR000884; TSP1.  
 DR InterPro: IPR002221; WAP.  
 DR Pfam: PF00014; Kunitz\_BPTI; 3.  
 DR PRINTS: PR00003; 4DISULPHCORE.  
 DR PRINTS: PR00759; BASICPTASE.  
 DR SMART: SM00409; IG; 3.  
 DR SMART: SM00408; IGC2; 2.  
 DR SMART: SM00410; IG-like; 1.  
 DR SMART: SM00131; KU; 3.  
 DR SMART: SM00209; TSP1; 7.  
 DR SMART: SM00217; WAP; 1.  
 DR PROSITE: PS50279; Bpti\_KUNITZ\_2; 3.  
 DR PROSITE: PS50092; TSP1; 3.  
 KW Matrix protein; Serine protease inhibitor; Signal.  
 FT SIGNAL 1  
 SQ SEQUENCE 2174 AA; 231936 MW; 03BF707952623120 CRC64;

Query Match 10.2%; Score 707; DB 5; Length 2174;  
 Best Local Similarity 29.2%; Pred. No. 4.1e-46;  
 Matches 220; Conservative 76; Mismatches 310; Indels 148; Gaps 29;

QY 545 PEQ--TYGDDG---WSSWTKFGSCSGGSGGVRNRSRSCNNPSPAYGGRPLGPMFEY 598  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 44 PESSVTPGGGNDPDEMTWSSPDSRTCGGGSVYQTECLRRDR-GEAVCSGSRRY 102  
 QY 599 QVNSEECPEGYEDFPAQCAK--RNSYVHQAQKSWPYEPDDDAKCELICOSADTG 656  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 103 FSCNTDQCEESDFFRAQOCSTRDROF--DGVFEWVY--TNAFPCLENCMPKGR 157  
 QY 657 DVFEMNOVYHDGTRCSYRDPYSCARGCVPYGCDEKVEGSMADKCGVCGGDNHSCRTV 716  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 158 FYRQREKVVYDTRCNDKD-LDYCVNGECMPYGCQDMLGSDAKEDKCRGGGDSSTCKTI 216  
 QY 717 KGTIGKASKQAGALKVQIPAGARIQIQALESKSHRSYVKNQYTSFTLNPKGEATSR 776  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 217 RNTITTKNLAPYNDLLPREGATNRIEETVSSNYLACRNH-SGHYYLNDWRIDRPE 275  
 QY 777 TTPAMGLEDAVED---AKESLKTSGPLPAIALALPPIEGGRSLAKKYIHEDL 832  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 276 PMEFANSMWNYQRKPMGFAPADOLTCSPISBLFTVMLVQER--NISLDEYISPEL 332  
 QY 833 LPLIGSNVNLLEMDTYEWALKSNAPCSKACGGGIGFTKYGCRRRRHHMQRHLCHDKK 892  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 333 SH-----SQDTHHTVTHOFGNACSASCGGSGNKKRYTCNNRTTLAEVPSLDCOKS 383  
 QY 893 RPKPIRRCONHPGSOPIVWTEEGACSGKLGAVGTGICLLPLSLNGTHKVPAPAKAC 952  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 384 KPYE-DQAGSTERCA-PHWVEGEMSKSGCGSDGFRNRSICERISSGSEHVEDAVC 441  
 QY 953 ---AGDRPEARPCLR--VPCPAQWRLGAMSGCSATCGEGIQOROVVCHTNNNSLGH---1004  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 442 LKEVGKMPAKQKQCNNDVKNCP-KYHLGPWTPCDKLGSGKQRTKTCFIEN--GHRKV 498  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 QY 1005 ---CEGDRPDTVOVCSLPACGNH-----ONSTRADVWELGTPEGQWVP 1046  
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 DB 499 LPEDDVEEKEPTEKSCLLTPCGVDMWISQMSGACNACQONTETRAI--CGNKBKGVY 556

QY 1047 QSGFLHPINKISSMCAAEPCGDRSVPCOMEV--LDRCYSIPGYHRLCCVSCITKASGPN 1104  
 Db 557 EE-----FCEPEVPTLSRPFCKSPKCEAQMWSSEMSKASAPC 592  
 QY 1105 P-----GPDGPTSPSPSTPGSPPLPGDPADAAEPKPGTSGEDHQ--HGRT 1152  
 Db 593 GKGVKSRIVICGEPDKT-----VTPADDSCKNKTKTESSEDDCEGEK 637  
 QY 1153 QLPALDTS-----SPTQHPFAPE--TPPGASMSISPTTPG 1188  
 Db 638 VCPGEWFTGPMGKCKSPGCGGGERVREVLCLSNCTKSVNCDKEEKLSEKNSACTEDE 697  
 QY 1189 GLPFWGWTQPTPYEDKGGEDLRHGGTSLPAD 1222  
 698 ILP--LSTDKPTEDDE---ECCDEDEGLIELSD 725

RESULT 8  
 QYVAV4 PRELIMINARY; PRT: 3060 AA.  
 ID QYVAV4: QYVAV3:  
 AC QYVAV4: QYVAV3:  
 DT 01-MAY-2001 (TEMBLrel. 13, Created)  
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE CG1540 PROTEIN.  
 GN PPN OR CG1540 OR CG18436.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Phrygoceta; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Phyloroidae; Drosophilidae; Drosophila.  
 NCBI\_TaxId=7227;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=BERKELEY;  
 RC MEDLINE=20196006; PubMed=10731132.  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Ramanides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,  
 Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,  
 Burks K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.D., Wei M.-H., Idegawa C.,  
 Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclik J.M.,  
 Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Splet E., Spradlin A.C., Stapleton M., Strong R., Sun E.,  
 Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang X., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).  
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 DR EMBL: AE003765; AAF56794.2; -;  
 DR EMBL: AE003765; AAF56795.2; -;  
 DR HSP: P12111; ZKNT.  
 DR FLYBase: FBgn0003137; Ppn.  
 DR Interpro: IPR000561; EGF-like.  
 DR Interpro: IPR003598; I9\_c2.  
 DR Interpro: IPR003006; Ig\_MHC.  
 DR Interpro: IPR002223; Kunitz\_BPTI.  
 DR Interpro: IPR000884; TSP1.  
 DR Interpro: IPR002221; WAP.  
 DR Pfam: PF00047; I9; 3.  
 DR Pfam: PF00014; Kunitz\_BPTI; 12.  
 DR Pfam: PF00090; TSP1; 5.  
 DR Pfam: PF00095; WAP; 1.  
 DR PRINTS: PR00759; BASICPTASE.  
 DR SMART: SM00408; ICG2; 3.  
 DR SMART: SM00131; K0; 12.  
 DR SMART: SM00209; TSP1; 7.  
 DR SMART: SM00217; WAP; 1.  
 DR PROSITE: PS00317; 4 DISULFIDE CORE; 1.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_1; 11.  
 DR PROSITE: PS00279; BPTI\_KUNITZ\_2; 12.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE: PS00092; TSP1; 3.  
 KW Alternative splicing; Immunoglobulin domain;  
 KW Serine protease inhibitor.  
 FT VARSPPLIC 2803 2803 L->SYSPV (IN SHORT ISOFORM).  
 FT VARSPPLIC 2844 2854 ENFKMEDSGI ->VASPLHPNAV (IN SHORT ISOFORM).  
 FT VARSPPLIC 2855 3060 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 3060 AA; 331579 MW; ACA31D3EE58C7C0 CRC64;

Query Match 10.2%; Score 707; DB 5; Length 3060;  
 Best local Similarity 29.2%; Pred. No. 6.8e-46;  
 Matches 220; Conservative 75; Mismatches 311; Indels 148; Gaps 29;

QY 545 PEO--TYGDGG---WSSMTKFGSCSRGCGGVSRSRSCNNPSPAYGAPCGPMPREY 598  
 Db 4 PESSVTPGEGNDPDEMTWSPSCSRFCGGVSYOTRELRDR-GEAVCGGSGRRY 62  
 QY 599 QVNSSECGTYEDBRAQCA--RNSYVHONAKHSVPEPDDAKCELTQSDATG 656  
 Db 63 FSCNTQDCPEESDPRADCSRFRQGF---DGYVEWVY--TNAVPCELNCPGGR 117  
 QY 657 DVYFNQVYHDGTGSRYPYSCVARGCEVPYCGDKVEGSMKADKCGYCGDMSHCRVY 716  
 Db 118 FYRQREKVVDDTRCNDK-LDYCVNGECMFGCDMMIGSDAKEDCKKCGGDSSTCKT 176  
 QY 717 KGTIGKASKQAGALKLVQIPAGARHIQLEALEKSPHSRVNQTGFIINPKKEATSR 776  
 Db 177 RWTITTKDLAPYNDLLLPBEGATNIRIEFVPSNVYLACNH-SGHYILNGWRIDPR 235  
 QY 777 TTGACLEHDAVED---AKSLKTSPPDEALAILALPTBEGPSSLAKKYVIEDL 832  
 Db 236 PWFANSMWNTQKRPDGFAPDOLTCSPISESLFIWLVQEK---NISDIYESIPESL 292  
 QY 833 LPLISNNVLLSEMTYMAKSNAPCSKAGGSIQFTYGCRRRRHHNVQHLCDHK 892  
 Db 293 SH-----SODTHVTHQFNAGASCAGGSGSRKVTYCNRRITLAEVNSLDDKS 343  
 QY 893 RPKPIRRCONPQSPQVWTEWAGSRSGKLGVTGRIGIQLPLTSLNTHKVMKAK 952  
 Db 344 KPYE-EQACGTEPCA-PRWVEGWSKSGSGGSPGNRBITERISSSEHVEEDAYC 401  
 QY 953 ---AGDRPARPCLR--VCPDAQWRLGANSQCATCSBGISQGRQVVCRTNANSLGH--- 1004  
 Db 402 LKEVGNKPKATQKQCNRPVKNCP-KYHILGPMTPDKILCGDKQTRKYTCFTEEN--GHRKV 458  
 QY 1005 -----CEGDRDVTQVQVSLPACGNH-----GNSTVRADVWELGTPBQWNP 1046



Db 459 LPEEDVEEKEPTEKSCILTPCEGVDTIISQMSGCNACQNTERTAL--CGNKEGKVP 516  
 QY 1047 QSGPLHPINKISSMCAAPRCIGDRSVCOMV--LDRYCSIPGYHRLCCVCICKASGPN 1104  
 Db 517 EE-----FCEPEYPTISRPCSKSPKCAQWFSSEMSKSCAPC 552  
 QY 1105 P-----GPDGPTSLPFPSTPGSPILPQPPADAAEPKPTGSEHQ--HGRAT 1152  
 Db 553 GAGVKSRIYICEFGDKT-----VTPADDSKCNKETKPESEODEGEEK 597  
 QY 1153 QLPALDTS-----SPGTHPRAPE--TPIGASKSISPTPG 1188  
 Db 598 VCPGEFTGPMGCKSPCGGGERVREVLQSLNGTSVNCDEKEKVEPLSEKCNSECTEDE 657  
 QY 1189 GLPMGWOTPTVPPEKKGQGEDLRHPTSLPAD 1222  
 Db 658 ILP--LTSTDKPIEDDE--EDCEDDGIELISD 685

UNT 9

PRELIMINARY: PRT: 2167 AA.

076840: Q22911:  
 076840: 01-NOV-1998 (Tremblrel. 08, Created)  
 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 C37C3.6 PROTEIN.  
 GN C37C3.6.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Feloderinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Geisel C., Bradshaw H.;  
 RT "The sequence of C. elegans cosmid C37C3."  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY  
 CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT  
 CC OF FORM B.  
 DR EMBL: U64857; AAC25868.1;  
 DR EMBL: U64857; AAC25868.1;  
 DR HSSP: P00981; 1DTR.  
 DR InterPro: IPR003598; IG\_C2.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR002223; Kunitz\_BPTI.  
 DR InterPro: IPR000884; TSP1.  
 DR Pfam: PF00047; Ig\_1.  
 DR Pfam: PF00014; Kunitz\_BPTI; 11.  
 RN Pfam: PF00090; tsp\_1; 6.  
 RN PRINTS: PR00759; BASICPTASE.  
 RN SMART: SM00408; IGC2; 1.  
 DR SMART: SM00131; KU; 11.  
 DR SMART: SM00209; TSP1; 7.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_1; 10.  
 DR PROSITE: PS50279; BPTI\_KUNITZ\_2; 11.  
 DR PROSITE: PS50092; TSP1; 4.  
 KW Alternative splicing; Hypothetical protein; Immunoglobulin domain;  
 KW Serine protease inhibitor.  
 FT VARSPIC 147 155 MISSING (IN ISOFORM A).  
 FT VARSPIC 1556 1558 KDD -> SKF (IN ISOFORM A).  
 FT VARSPIC 1559 2167 MISSING (IN ISOFORM A).  
 FT SEQUENCE 2167 AA; 23759 MW; 96274786D52E3639 CRC64;

Query Match 10.1%; Score 699.5; DB 5; Length 2167;  
 Best Local Similarity 30.0%; Pred. No. 1.6e-45;  
 Matches 179; Conservative 86; Mismatches 225; Indels 107; Gaps 26;

QY 551 QDGGSWTFRGSCGSGGVRSRSCNNPRAVGRCLCPMFYQVCNSEEC-PGT 609

Db 75 EPTNMGWVWVNECNRSRSCGGVQLEKRCSSG-----DCTGASVYISCLNINCESGT 126  
 QY 610 YEDFPAQCAKRSYVYHONAKSWVPEYEDDDAQCELIQASADVDYV--MMQVYND 667  
 Db 127 -DFRAEOCSKFDEALDGN-YHKWTPYK--GKNKCELVCK-PESGNFYVWADKVY-D 178  
 QY 668 GTRCSYDPTSVYCARGVYVCGDCKEYGSWMKADKCGVCGDMSHCRKVTGLGKASKQA 727  
 Db 179 GTRKDSKS-NDICVDGECFLVCGCDKLGSSLKFKDKCKGCKDGDOSTCKTLEGREDENLSP 237  
 QY 728 GALKVQIPAGARHIQLEALEKSPHSVYKQVYGS--FLNPKGEATSRFTTAMG--L 783  
 Db 238 GYHDIKILPGATQNIKIQEARSTNNLALKN--GSDHFTYINGNGLIYQVEYEGVGTIF 294  
 QY 784 EWEDAVDAKESLKTSPLEALAILALPTEGPPSSLAQKVIHEDLLPLGSSNVLL 843  
 Db 295 YVDDA--EPETLSAQSPLEELTV-ALLFRKGSRDALTYE-----SIFL 337  
 QY 844 EEMDTYEMALKSNAPCSKACGGGIGFTKYGCRRRRDHMYRHLCDHKRPK-IRRCGN 902  
 Db 338 EEEVDYMKKEDNMPICVSCGKGYOTNLICIDGKNGRVEDDLCEBNATKPEFEKCE 397  
 QY 903 QHPCSQPVWVTEEMGACSRSCGLVOTRGICLLPLSNQTHVMRAKACAGDREPARP 962  
 Db 398 TVDC-EAEWFTGDMWESSSTCGDGOQYRVYCHQVYFANGRVYEDGNTVERPPVKOT 456  
 QY 963 CLRPVCPAOWMLGMSQCSATCGEGIOQROVVCRTN-----ANSIGHEGDRPPT 1012  
 Db 457 CNRFACP-EMWAGPMSACSEKCGDAFYRSVTCSEKEGEGKLIADAPADQEKFPD 515  
 QY 1013 VOVCSLPACGGNHONSTVRADVMEL-----GTREGQWVPOSGPLHPINKI----- 1057  
 Db 516 ERTCNLGPCEG---LTFVIGEMNICTRCNDTEETREVTCKDSQGRAYPEKCLVNSTE 571  
 QY 1058 ---SSMCAAP-----CTGD-----RSVFCQ-----MEVLDK-YC 1083  
 Db 572 IPTDIRSCATQPCPEYEWVSEMSKCTTECGHGKTRVICALIHONGLEIVDGHG 628

RESULT 10  
 095428 PRELIMINARY: PRT: 1235 AA.  
 AC 095428:  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE HYPOTHETICAL 133.5 KDA PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RA Rowen L., Madan A., Qln S., Abbasi N., Dors M., Ratcliffe A.,  
 RA Madan A., Dickhoff R., Shaffer T., James R., Lasly S., Hood L.;  
 RT "Complete sequence of the gene for presenilin 1."  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF109907; AAC97963.1;  
 DR HSSP: P12111; 2KNT.  
 DR InterPro: IPR003598; IG\_C2.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR002223; Kunitz\_BPTI.  
 DR InterPro: IPR000884; TSP1.  
 DR Pfam: PF00047; Ig\_3.  
 DR Pfam: PF00014; Kunitz\_BPTI; 1.  
 DR Pfam: PF00090; tsp\_1; 5.  
 DR PRINTS: PR00759; BASICPTASE.  
 DR SMART: SM00408; IGC2; 3.  
 DR SMART: SM00131; KU; 1.  
 DR SMART: SM00209; TSP1; 5.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE: PS50279; BPTI\_KUNITZ\_2; 2.  
 DR PROSITE: PS50092; TSP1; 4.







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2002, 00:04:09 ; Search time 88.81 Seconds

(without alignments)  
409,064 Million cell updates/sec

Title: US-09-938-330-22\_COPY\_1040\_1249  
Perfect score: 1206  
Sequence: 1 PEGQWVPSGSLPINKISS.....PTGFTLCVLPDSQLAGHT 210

String table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Search: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_ARCHAEA:  
2: SP\_BACTERIA:  
3: SP\_FUNGI:  
4: SP\_HUMAN:  
5: SP\_INVERTEBRATE:  
6: SP\_MAMMAL:  
7: SP\_MNC:  
8: SP\_ORGANELLE:  
9: SP\_PHAGE:  
10: SP\_PLANT:  
11: SP\_RECENT:  
12: SP\_VIRUS:  
13: SP\_VERTEBRATE:  
14: SP\_UNCLASSIFIED:  
15: SP\_VIRUS:  
16: SP\_BACTERIAP:  
17: SP\_ARCHAEP:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	204	16.9	364	4	096AY5	096AY5 homo sapien
2	180.5	15.0	839	10	09SN46	09SN46 arabidopsis
3	175.5	14.6	585	12	041935	041935 murid herpe
4	163	13.5	345	5	095W2	095W2 meloidogyne
5	160	13.3	319	5	017038	017038 caenorhabdi
6	160	13.3	724	11	099LH3	099LH3 mus musculu
7	158	13.1	651	4	09NX19	09NX19 homo sapien
8	158	13.1	845	4	09HB5	09HB5 homo sapien
9	158	13.1	845	4	09HAU3	09HAU3 homo sapien
10	156.5	13.0	309	5	018751	018751 caenorhabdi
11	156	12.9	845	4	09HB8	09HB8 mus musculu
12	155.5	12.9	768	11	099N39	099N39 mus musculu
13	154	12.8	577	10	09STP1	09STP1 arabidopsis
14	153	12.7	408	12	09J2H0	09J2H0 macaca mula
15	151.5	12.6	305	5	025467	025467 meloidogyne
16	151	12.5	448	12	09WRM2	09WRM2 macaca mula

17	150.5	12.5	316	5	093207	093207 caenorhabdi
18	149.5	12.4	301	5	019763	019763 caenorhabdi
19	149.5	12.4	995	4	09Y2M4	09Y2M4 homo sapien
20	149.5	12.4	1012	4	043393	043393 homo sapien
21	149.5	12.4	1012	4	075359	075359 homo sapien
22	149.5	12.4	1268	4	075046	075046 homo sapien
23	149.5	12.4	1566	4	09P2R6	09P2R6 homo sapien
24	149	12.3	325	5	017036	017036 caenorhabdi
25	148.5	12.3	555	10	09P0C6	09P0C6 chlamydomon
26	148	12.3	676	6	095J7C9	095J7C9 sus scrofa
27	146.5	12.1	846	11	099N40	099N40 mus musculu
28	146	12.1	290	5	022260	022260 caenorhabdi
29	145.5	12.1	299	5	019218	019218 caenorhabdi
30	145.5	12.1	316	5	093210	093210 caenorhabdi
31	145.5	12.1	330	10	049870	049870 hordew vul
32	145.5	12.1	566	6	095J7D1	095J7D1 sus scrofa
33	145.5	12.0	297	4	016038	016038 homo sapien
34	145	12.0	297	5	023263	023263 caenorhabdi
35	145	12.0	299	5	020087	020087 caenorhabdi
36	145	12.0	358	4	008805	008805 homo sapien
37	145	12.0	1180	5	09VRM2	09VRM2 drosophila
38	144.5	12.0	313	5	020922	020922 caenorhabdi
39	144.5	12.0	356	5	020921	020921 caenorhabdi
40	144.5	12.0	511	6	095J7D0	095J7D0 sus scrofa
41	144.5	12.0	888	13	090796	090796 gallus gall
42	144	11.9	362	5	09VC23	09VC23 drosophila
43	143.5	11.9	196	5	09GCT8	09GCT8 heterodera
44	143.5	11.9	642	10	09LD34	09LD34 cryptospori
45	143.5	11.9	862	11	09JIK1	09JIK1 rattus norv

## ALIGNMENTS

RESULT 1  
ID 096AY5 PRELIMINARY: PRT: 364 AA.  
AC 096AY5;  
DT 01-DEC-2001 (TREMBLrel, 19, Created)  
DT 01-DEC-2001 (TREMBLrel, 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel, 19, Last annotation update)  
DE HYPOTHETICAL 40.2 KDA PROTEIN (FRAGMENT).  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA, AND CHORIOCARCINOMA;  
RA Strausberg R.;  
Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC016451; AAI16451.1;  
KW Hypothetical protein.  
FT NON\_TER  
KW NON\_TER  
SQ SEQUENCE 364 AA: 40210 MW: A3D54D967D67FD30 CRC64;

Query Match 16.9%; Score 204; DB 4; Length 364;  
Best Local Similarity 38.7%; Pred. No. 2.5e-08;  
Matches 58; Conservative 12; Mismatches 36; Indels 44; Gaps 8;  
QY 12 LHPINKISTEPCDDRSVFCQMEVLDYRCSIPGHRLCVCYCIKKASG----- 60  
DB 171 LPPCN-----DEPCIDCKSTFCQMEVLAIRCSIPGHRLCVCYCIKKASG----- 60  
QY 61 -----PNEGPPPG-----PTSLPPF--STPG-----SPLPQOPADAAEPGKPTG 100  
DB 227 THDDVYNSPDLPRSLVPTSLVPRHSETPAKKMSLSISVGGPNAYA-AFRPNKSPDS 205  
QY 101 SEDRHGGRATQIPGALDT-----SSPECTQ 124  
DB 286 ANLRQ--RSAQAGSKTVRLVTPSPPTK 313

DT	ID	AC	Q9SN46	RESULT	2
DT	01-MAY-2000	(TEMBLrel. 13, Created)			
DT	01-MAY-2000	(TEMBLrel. 13, Last sequence update)			
DT	01-DEC-2001	(TEMBLrel. 19, Last annotation update)			
DE	EXTENSIN-LIKE PROTEIN.				
GN	F28A21.80 OR AT4G18670.				
OC	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;				
OC	Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;				
OC	eurosid II: Brassicales; Brassicaceae; Arabidopsi.				
OX	NCBI_Taxid=3702;				
RN	SEQUENCE FROM N.A.				
RA	Beyan M., Mueller M.W., Muehleln A., Felber R., Bancroft I.,				
RA	Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;				
RA	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.				
RL	SEQUENCE FROM N.A.				
RN	EU Arabidopsis sequencing project;				
RN	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.				
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
RN	SEQUENCE FROM N.A.				
RA	Mayer K.F.X.;				
RA	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
RL	SEQUENCE FROM N.A.				
RA	EU Arabidopsis sequencing project;				
RA	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AL035526; CAB37452.1; -				
DR	EMBL: AL161549; CAB7869.1; -				
DR	InterPro: IPR002951; Atrophin.				
DR	InterPro: IPR001611; LRR.				
DR	InterPro: IPR003592; LRR_out.				
DR	InterPro: IPR002965; P_Rich_extensn.				
DR	Pfam: PF00560; LRR; 6.				
DR	PRINTS: PR01222; ATROPHIN.				
DR	PRINTS: PR01217; PRICHEXTENS.				
DR	SMART: SM00370; LRR: 4.				
DR	SEQUENCE 839 AA; 90243 MW; 0F0F04F22D3D9789 CRC64;				
QY	Query Match	15.0%; Score 180.5; DB 10; Length 839;			
QY	Best Local Similarity	29.2%; Pred. No. 3.6e-06;			
QY	Matches	59; Conservative 10; Mismatches 64; Indels 69; Gaps 9.			
QY	1 PEGQWPGGPIHPINKISTEPCGDRNVFCOMEVLDRCSPGYNRLCVCSCIKKASG 60				
QY	411 PSEPPTSPGSGSPSSISPSPTIT-----VP-----SPTT 442				
QY	61 PNGG--PDPGTSLPPEST--PGSFLPQPDADAEPPKGTSGEDHNGRATOLPGA 115				
QY	443 PSPGGSPSPSIYVSPSTTSPGSPSPSTPTPTPGSGSPSSPT----- 486				
QY	116 LDTSSGCTQHPAPETPTTGASWISPTT--PGRLWGMGTQPTP---VPEDKGGPGEDL 170				
QY	487 --TPRTGSGPSSPTPTPTGSGSPSSPTTPSPGSGSPSSISPSPTTYVSPSTP---- 540				
QY	171 RHFGTSLPADLGRPRECHPT 192				
QY	541 ----TS-----PGSPSPSSPT 553				
DT	01-JAN-1998	(TEMBLrel. 05, Created)			
DT	01-JAN-1998	(TEMBLrel. 05, Last sequence update)			

DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE HYPOTHETICAL 60.2 KDA PROTEIN.  
GN GAMAHV.M6.  
OS murid herpesvirus 4.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae.  
OX NCRL\_TaxID=33708;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WUMS;  
RX MEDLINE=97366649; PubMed=9223479;  
RA Virgin H.W. IV, Latelle P., Wamsley P., Hallsworth K., Weck K.E.,  
RA Dal Canto A.J., Speck S.H.;  
RT "Complete sequence and genomic analysis of murine gammaherpesvirus  
68.";  
RL J. Virol. 71:5894-5904(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WUMS;  
RA Latelle P., Wamsley P., Waterston R.H.;  
RL Submitted (Apr-1997) to the EMBL/Genbank/DBJ databases.  
DR EMBL: U97553; AAB6392.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 585 AA; 60160 MW; 85610A80C34827D CRC64;

Query Match		14.6%;	Score 17.5.5;	DB 12;	Length 585;
Best Local Similarity		30.0%;	Pred. No. 6.2e-06;		
Matches 63; Conservative		12;	Mismatches 72;	Indels 63;	Gaps 12
Oy	40 YCSIPGYHNLCCVSC-----IKKAGPNP--GDDGPSTLPPFTPGSP---	81			
Dd	55 FSLSSGFMCSSCTCADDRFGFPGRAGARIDPLDGPLPSMGPDPRPPRPPELPGSPTSP	114			
Oy	82 -----LPGR-----ODPADAEF---PGKPTSSEDHQHG-RATQLFGALDTSS	120			
Dd	115 ABSRAGARIIDLGLPGLPSMGDPDRPPRRPPRPPELPGSGTSPAASRAARIDPLDGPLPSMG	174			
Oy	121 PGTO-----HPAPETPIPGASWSISPTTPGGIPL-WGTQTPTPVVEDKGQCGEDL-	170			
Dd	175 PDPRPPRPPELPGGSPTSPAASRAARIDLDGLPLPSMGPDPRPPRPPELG-PGSPTS	233			
Oy	171 ---RHGTSILPADLPG-----HPPER 188				
Dd	234 PAPSRAGARIP-DLPGDLPSMGPDPRPPRP 262				
RESULT	4				
O95WY2	PRELIMINARY;	PRT:	345 AA.		
AC	O95WY2;				
DT	01-DDC-2001 (TREMBLrel. 19, Created)				
DT	01-DDC-2001 (TREMBLrel. 19, Last sequence update)				
DT	01-DDC-2001 (TREMBLrel. 19, Last annotation update)				
DE	COLLAGEN.				
GN	COL-5.				
OS	Meloidogyne javanica (Root-knot nematode).				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;				
OC	Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.				
NB	NCBI_TaxID=6303;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=21425611; PubMed=11534644;				
RA	Iltu T., Koltai H., Chejanovsky N., Spiegel Y.;				
RT	"Isolation of a novel collagen gene (Mj-col-5) in Meloidogyne javanica				
RL	and analysis of its expression pattern.";				
J. Parasitol.	87:801-807(2001).				
EMBL:	AJ289026; AAK83075.1; "				
SQ	SEQUENCE 345 AA; 35524 MW; C4034D717DA48BD0 CRC64;				
Query Match		13.5%;	Score 163;	DB 5;	Length 345;
Best Local Similarity		27.8%;	Pred. NO. 3.4e-05;		



RT "NEDO human cDNA sequencing project."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK000226; BAA91021.1; -  
 DR InterPro: IPR002126; Cadherin.  
 DR SMART: SM00112; CA; 1.  
 DR PROSITE: PS50268; CADHERIN\_2; 3.  
 DR SEQUENCE 651 AA; 67379 MW; AE6D4984FD69175C CRC64;

Query Match 13.1%; Score 158; DB 4; Length 651;  
 Best Local Similarity 26.1%; Pred. No. 0.00015;  
 Matches 63; Conservative 5; Mismatches 69; Indels 104; Gaps 11;

QY 53 SCIRKASGPNPDPDPPTSLPPFS-----TPG-----SPLPGQDPA 89  
 DB 326 STTSSGEGTGHPPPGTTLRPPTSTPGSGPAENSTSHQATPGDTAQTPEKPGTSQPM 385  
 QY 90 DAAPPGKPTGSEDDHGRATOLPGALDTSPPGTOHPFAP----- 129  
 DB 386 ---PPGVGTST---SHQATPGSGTGTVPETGTSQPMPPSMGTSSTSHQATPGGTAQT 438  
 QY 130 -----ETPIP---GASWSISPTTPPGGLPWGWTOTPTP-----VPEDKQPGEDL 170  
 DB 439 PEAGTSQPMPPGKGTSTSHQPTTPG---GTAQTPEPGTSQPMPLSKTSPSSGGGSEBK 495  
 QY 171 R-----HPGTSPLADLGRPREPCHPTGFTLCVL 200  
 DB 496 RFSVYDMAALGVGLALLLALLGLAVLVKHKYGRPLKC-CSGKAPDP-OPQGFNDQAF 553  
 QY 201 P 201  
 DB 554 P 554

RESULT 8  
 Q9HBB5 ID Q9HBB5 PRELIMINARY; PRT; 845 AA.

AC Q9HBB5: 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MCDHL-FL.  
 GN MCDHL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Paris M.J., Williams B.R.G.;  
 PM "MCDHL, a Novel cDNA with Mutin and Cadherin-like Domains,"  
 submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 EMBL: AF558676; AAG16733.1; -  
 InterPro: IPR002126; Cadherin.  
 SMART: SM00112; CA; 3.  
 DR PROSITE: PS00232; CADHERIN\_1; 1.  
 DR PROSITE: PS50268; CADHERIN\_2; 4.  
 DR SEQUENCE 845 AA; 88156 MW; 10E654279CE30ECE CRC64;

Query Match 13.1%; Score 158; DB 4; Length 845;  
 Best Local Similarity 26.1%; Pred. No. 0.00019;  
 Matches 63; Conservative 5; Mismatches 69; Indels 104; Gaps 11;

QY 53 SCIRKASGPNPDPDPPTSLPPFS-----TPG-----SPLPGQDPA 89  
 DB 492 STTSSGEGTGHPPPGTTLRPPTSTPGSGPAENSTSHQATPGDTAQTPEKPGTSQPM 551  
 QY 90 DAAPPGKPTGSEDDHGRATOLPGALDTSPPGTOHPFAP----- 129  
 DB 552 ---PPGVGTST---SHQATPGSGTGTVPETGTSQPMPPSMGTSSTSHQATPGGTAQT 604  
 QY 130 -----ETPIP---GASWSISPTTPPGGLPWGWTOTPTP-----VPEDKQPGEDL 170

DB 605 PEAGTSQPMPPGKGTSTSHQPTTPG---GTAQTPEPGTSQPMPLSKTSPSSGGGSEBK 661  
 QY 171 R-----HPGTSPLADLGRPREPCHPTGFTLCVL 200  
 DB 662 RFSVYDMAALGVGLALLLALLGLAVLVKHKYGRPLKC-CSGKAPDP-OPQGFNDQAF 719  
 QY 201 P 201  
 DB 720 P 720

RESULT 9  
 Q9HAU3 ID Q9HAU3 PRELIMINARY; PRT; 845 AA.

AC Q9HAU3: 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE MU-PROTOCADHERIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tissue-KIDNEY;  
 RC Soileman A., Krieger S., Haase A., Hantusch B.;  
 RL "Cloning and characterization of human mu-protocadherin,"  
 submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF301909; AAG33495.1; -  
 DR InterPro: IPR002126; Cadherin.  
 DR SMART: SM00112; CA; 3.  
 DR PROSITE: PS00232; CADHERIN\_1; 1.  
 DR PROSITE: PS50268; CADHERIN\_2; 4.  
 DR SEQUENCE 845 AA; 88171 MW; D12C11C1E3E11680 CRC64;

Query Match 13.1%; Score 158; DB 4; Length 845;  
 Best Local Similarity 26.1%; Pred. No. 0.00019;  
 Matches 63; Conservative 5; Mismatches 69; Indels 104; Gaps 11;

QY 53 SCIRKASGPNPDPDPPTSLPPFS-----TPG-----SPLPGQDPA 89  
 DB 492 STTSSGEGTGHPPPGTTLRPPTSTPGSGPAENSTSHQATPGDTAQTPEKPGTSQPM 551  
 QY 90 DAAPPGKPTGSEDDHGRATOLPGALDTSPPGTOHPFAP----- 129  
 DB 552 ---PPGVGTST---SHQATPGSGTGTVPETGTSQPMPPSMGTSSTSHQATPGGTAQT 604  
 QY 130 -----ETPIP---GASWSISPTTPPGGLPWGWTOTPTP-----VPEDKQPGEDL 170  
 DB 605 PEAGTSQPMPPGKGTSTSHQPTTPG---GTAQTPEPGTSQPMPLSKTSPSSGGGSEBK 661  
 QY 171 R-----HPGTSPLADLGRPREPCHPTGFTLCVL 200  
 DB 662 RFSVYDMAALGVGLALLLALLGLAVLVKHKYGRPLKC-CSGKAPDP-OPQGFNDQAF 719  
 QY 201 P 201  
 DB 720 P 720

RESULT 10  
 Q18751 ID Q18751 PRELIMINARY; PRT; 309 AA.

AC Q18751: 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE C50F7.5 PROTEIN.  
 GN C50F7.5  
 OS Caenorhabditis elegans.

[illegible]

DR PROSITE; PS50268; CADHERIN\_2; 4.  
SQ SEQUENCE 845 AA; 88097 MW; 6CD0E3985391F59F CRC64;

		12.9%	Score 156;	DB 4;	Length 845;	
		Best Local Similarity 26.1%;	Pred. No. 0.00029;			
		Matches 63;	Conservative 4;	Mismatches 70;	Indels 104;	Gaps 11;
QY	53	SCIKKASGPNCGDDEGPTSLPPFS-----TPG-----SPYGGQDDA	89			
Db	492	STTSSSGGGPHPPSPGCTTLRPPSTSTPGSGSGAENKSTSHOPATGGGDTAQTPEKGTSSQPM	551			
QY	90	DAEPEPKPTGSEDHQHGKATOLPCALOTSSPGTQHPAP-----	129			
Db	552	-----PGVGSTST--SHOPATPSGGTAQTPEPKGTSSQPMPSMGSTSHOPATGGGTAQT	604			
QY	130	-----ETPLP---GASMSISPTTGGCLPWCMTQTPP-----VPEDKGAGCEDL	170			
Db	605	PEAGTSPQMPMPGMGTSTSHOPTTPOG---GTAQGTPEPGTSSQPMWLSKSTPSSGGSGSEDK	661			
QY	171	R-----HPCGSLPADLPGRHPPECPHTGTFLLCVL	200			
Db	662	RFSVVDMAALGGVIALALLLLALLGLAVLVYHHNHYPRKLC-CSGKAPEP-QPGSGFDMAQFL	719			
QY	201	P 201				
Db	720	P 720				

RESULT	12			
099N39				
ID	099N39	PRELIMINARY;	PRT;	768 AA.
AC	099N39			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DD	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	NADRLIN2.			
GN	NMADRINZ.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6; TISSUE=BRAIN;			
RA	Kobayashi Y., Furuta B., Itoh K., Umega M.;			
RT	"Mus musculus homolog of rat Nadrin."			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AB060554; BAB43663.1; "			
DR	HSPF; 007960; IKG.			
DR	InterPro: IPR002965; P_rich_extensn.			
DR	InterPro: IPR000198; RhogCAP.			
DR	Pfam: PF006620; RhogCAP. 1.			
DR	PRINTS: PR01217; PRICHEXTENS.			
DR	SMART: SM00324; RhogAP. 1.			
DR	SEQUENCE 768 AA; 84309 MW; C1CFLIC73CA7044F CRC64;			

Query Match	12.9%	Score 155.5	DB 11	Length 768
Best Local Similarity	25.5%	Pred. No. 0.00028		
Matches 62	Conservative 20	Mismatches 92	Indels 69	Gaps 9

QY	2	EGWVPSGSLHAI	INKISSTPCTGDSVFCMEYLDRCYSIPGHRHLCVCSCIKASG	61
Db	489	EGDLVKKESPPK	KDSDSAVPAAGRDS--NMTTVPAQATGCGNSQLSVSPHSAAG	546
QY	62	NP-----	GDPGTSLP-----FSTPG--SPLRGDPADAAEPGKGTSGS	102
		:	:	:
Db	547	SPHTLRRAVK	KPARAPRKPGNLPRPGHGGSSPGTGSTPKRSASR---SPQQQDQDQDQD	603
QY	103	DHNGRATQL	PGALDTSS-----PGQHMFATETIPGASMSISPTTPGCL	148
		:	:	:
Db	604	QDQDQDQDQDQD	PGARRCSSLPIQASHPRPQRPRLGDSGGDEGPPRPQRPTRPSPRP	663





DR InterPro: IPR002486; COL\_cuticle\_N.  
 DR Pfam: PF01391; Collagen; 2.  
 DR Pfam: PF01484; COL\_cuticle\_N; 1.  
 SO SEQUENCE 305 AA; 30586 MW; F24770C129694718 CRC64;

Query Match 12.6%; Score 151.5; DB 5; Length 305;  
 Best Local Similarity 31.3%; Pred. No. 0.00023;

Matches 71; Conservative 12; Mismatches 59; Indels 85; Gaps 17;

QY 7 PQSGPLPINKISTEPCTGDRSVFCOMENVLDRCSTIPGHRLCCVCSICKASGPNCGPD 66  
 DB 71 PRSNPF-----QSLYRQKRSLPDVC-----ICQPLEINAPRP-RGP- 106  
 67 PGPTSLP-----PFS--TPGSPLEPCPD-----PADAAEPGK--PTGSEDHQH 106  
 107 PGPPEQPEHPEQPEHVGQPGSPGAPPCPLPDQACORCAPAGAGTGTGKGGPAG-QPGP 165  
 107 GRATQLPGALDTSS-----PGTQHPFAPETPT-----PGASWSISPTTGGGLPMG 151  
 DB 166 GR-----PGAPKSSGAGAPPGAPGCGPPGAGKNGGPGCGQPKNGVSHPTTIPG--PKG 219  
 QY 152 WTQPTPTVPEDKGGPGGEDLRHPTGSLPADLPGR-----PRECHPTG 193  
 DB 220 PGGSP-----GQPKR---PG---PAGVAGKTGPGEGPPGVPAG 252

RESULT 16

Q9WRM2 PRELIMINARY; PRT; 448 AA.

AC Q9WRM2; 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE LATENT NUCLEAR ANTIGEN.  
 OS Macaca mulatta rhadinovirus 17577.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Rhadinovirus.  
 NCBI\_TaxID=83534;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99174001; PubMed=10074154;  
 RA Searles R.P., Berguam E.P., Axthelm M.K., Wong S.W.;  
 RT "Sequence and genomic analysis of a rhesus macaque rhadinovirus with  
 RT similarity to Kaposi's sarcoma-associated Herpesvirus/Human  
 RT herpesvirus 8.";  
 RT J. Virol. 73:3040-3053(1999).  
 [2]  
 SEQUENCE FROM N.A.  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF083501; AAD21406.1;  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR InterPro: IPR000531; TonB\_boxC.  
 DR PRINTS: PR01217; PRICHEXTENSN.  
 DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; UNKNOWN\_1.  
 SO SEQUENCE 448 AA; 47459 MW; E69BE03FEE4A974F CRC64;

Query Match 12.5%; Score 151; DB 12; Length 448;  
 Best Local Similarity 27.6%; Pred. No. 0.00036;

Matches 63; Conservative 10; Mismatches 71; Indels 84; Gaps 15;

QY 1 PEGQWVPSGFLP-----INKISTEPCTGDR-SVFCOMENVLDRCSTIPGHRLCCVSCI 55  
 DB 132 PASHSPSPSPSPHPGHSAPPSHPSPNOQPSSTLQ-----PSH----- 173  
 QY 56 KKASGPNCGPDGPTSLPPTSTPGSP-----IPGQDPADAAEPGKPTGSEDHQH 108  
 DB 174 ---DSBP--PEP--PTSLPPTDSGPPGSPPTSSPPGSPGSDPPTGQ----- 222  
 QY 109 ATQLPGALDTSSTPGTQ-----HPAPF---TPRG-----ASWSISPTTGGGLP- 149

DB 223 -----APSPNTQAVSHTDHPGSPRPGPPPGHTSHSYVGWG--PPTRAGVPC 272  
 QY 150 -----MGWTPPTPVE-DKQPGEDLRHPTGSLPADLPGRPREPCHP 191  
 DB 273 LRLRCTSHNSHDEAPEROEGEERQD-----PARPPRPPR 312

RESULT 17

Q93207 PRELIMINARY; PRT; 316 AA.

AC Q93207; 01-FEB-1997 (TREMblrel. 02, Created)  
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE C15A11.6 PROTEIN.  
 GN C15A11.6.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gardner A.E.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL: Z79694; CAB01958.1;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR002486; COL\_cuticle\_N.  
 DR Pfam: PF01391; Collagen; 3.  
 DR Pfam: PF01484; COL\_cuticle\_N; 1.  
 SO SEQUENCE 316 AA; 31232 MW; 1FC26A3770E9A6D5 CRC64;

Query Match 12.5%; Score 150.5; DB 5; Length 316;  
 Best Local Similarity 30.9%; Pred. No. 0.00028;  
 Matches 56; Conservative 13; Mismatches 73; Indels 39; Gaps 12;

QY 44 PGYN-RLCVCSTCKKASGNPBP-DPGPTSLPPTSTPGSL-----PGQDPADAAEP 94  
 DB 125 PSHHQKQECIKCPQGLPGAGAPGQPPRG--PNCNPGAPAGGSGGCGPPGPGAGSAGS 182  
 QY 95 PGKP-----TGSEDHQHGATQLPGALDTSSPTQHPFAPETP--IPGASWSISPTTGG 148  
 DB 183 PGAGAPGPNRPGSPGKSGRGRLPG--PSGAPGSGP--PGAPGPGGGMNPPGAPPG- 237  
 QY 149 PNGWTPPTVPEDKGGPGGEDLRHPTGSLPADLPGRPREPCHPTGTTFLCVLPDSTQRLG 208  
 DB 238 PAG-----PNCQPG---HPGQD---GPGAPGNDGTPGSDAAYCPCPTRSSVLR 280  
 QY 209 H 209  
 DB 281 H 281

RESULT 18

Q19763 PRELIMINARY; PRT; 301 AA.

AC Q19763; 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE F23H12.4 PROTEIN.  
 GN F23H12.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NCBI\_TaxID=6239;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.K.,  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 investigating biology."  
 RL Science 282:2012-2018(1998).  
 DR EMBL: Z74472; CA98942.1; -  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR002486; COL\_cuticle\_N.  
 DR Pfam: PF01391; Collagen; 2.  
 DR Pfam: PF01484; COL\_cuticle\_N; 1.  
 SQ SEQUENCE 301 AA; 29177 MW; 4A75B65540E50360 CRC64;

Query Match 12.4%; Score 149.5; DB 5; Length 301;  
 Best Local Similarity 34.9%; Pred. No. 0.00032;  
 Matches 59; Conservative 6; Mismatches 69; Indels 35; Gaps 13;

QY 51 CVCSTKASGPNP-----DPGPTSLP---PFSTPGSLP-GPDDPADAEPPKPTG 100  
 ID 09Y2M4 PRELIMINARY; PRT; 995 AA.  
 AC 09Y2M4;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE ATROPHIN-RELATED PROTEIN ARP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20195630; PubMed=10729226;  
 RA Anker L.C., Bauser A., Corvi R., Dhlmann S., Praml C., Cavenee W.K.,  
 RT Schwab M., Hampton G.M.,  
 RT "Identification and characterization of novel genes located at the  
 t(1;15)(p36.2;q24) translocation breakpoint in the neuroblastoma cell  
 line NGP."  
 RL Genomics 64:195-202(2000).  
 DR EMBL: AF118275; AAD27584.1; -  
 DR InterPro: IPR002951; Atrophin.  
 DR InterPro: IPR002965; P-rich\_extensn.  
 DR PRINTS: PRO1222; ATROPHIN.  
 DR PRINTS: PRO1217; PRICHEXTENSIN.  
 SQ SEQUENCE 995 AA; 107088 MW; 888B827A44BFB04 CRC64;

RESULT 19  
 QY 160 PEDKGGEDLRHPTSLRPLDGRPERCHPTGTFILCVLPDSOLRG 208  
 DB 247 PGPAGGPPD-GPPVDGSGSPPG-PGPAGTGEKGC--PKYCALDG 291

Query Match 12.4%; Score 149.5; DB 4; Length 995;  
 Best Local Similarity 31.2%; Pred. No. 0.001;  
 Matches 50; Conservative 9; Mismatches 74; Indels 27; Gaps 6;

QY 64 GPDGPTSLPFGTSGSLPGRPDADAEPPKPTGSEDHQHRATQLPGALDTSSTG 123  
 DB 175 GVPAPSSAPP-GTPQLTPGPPTPSATVPPQGSPTASQ-----APNQPAPTAIVPHT 227

QY 124 Q-----HPFAPETPIPGASWSISP-----TTPGGLPWCWTQTTPVPEDKGGPGE 169  
 DB 228 HIQAPALHPQRPPSPHPPHPSHPRLPLTGSAQGPSASHPRL-HQGGPPGHS 286  
 QY 170 -----LRHPTSLPADLGRPPPECHPTGTTLCVLPDRS 204  
 DB 287 QAGPLQHHPGPPOPFGLEPPQASQOAPLGTSPAAAYPHTS 326

RESULT 20  
 QY 043393 PRELIMINARY; PRT; 1012 AA.  
 ID 043393;  
 AC 043393;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JAN-1999 (TREMBlrel. 09, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE ATROPHIN-1 RELATED PROTEIN.  
 GN DRPLA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Xia J.-H., Liu C.-Y., Ruan Q.-G., Wang D.-A., Deng H.-X.;  
 RT "Cloning and localization of Human Atrophin-1 (DRPLA) related gene."  
 RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF016005; AAC31120.1; -  
 SQ SEQUENCE 1012 AA; 108967 MW; 4DMA8872AB78866 CRC64;

Query Match 12.4%; Score 149.5; DB 4; Length 1012;  
 Best Local Similarity 31.2%; Pred. No. 0.001;  
 Matches 50; Conservative 9; Mismatches 74; Indels 27; Gaps 6;  
 QY 64 GPDGPTSLPFGTSGSLPGRPDADAEPPKPTGSEDHQHRATQLPGALDTSSTG 123  
 DB 192 GVPAPSSAPP-GTPQLTPGPPTPSATVPPQGSPTASQ-----APNQPAPTAIVPHT 244  
 QY 124 Q-----HPFAPETPIPGASWSISP-----TTPGGLPWCWTQTTPVPEDKGGPGE 169  
 DB 245 HIQAPALHPQRPPSPHPPHPSHPRLPLTGSAQGPSASHPRL-HQGGPPGHS 303  
 QY 170 -----LRHPTSLPADLGRPPPECHPTGTTLCVLPDRS 204  
 DB 304 QAGPLQHHPGPPOPFGLEPPQASQOAPLGTSPAAAYPHTS 343

Search completed: August 29, 2002, 00:11:10  
 Job time: 421 sec

Thu Aug 29 10:46:29 2002

us-09-938-330-22\_copy\_1040\_1249.rpt



Result	Query	Match	Length	DB	ID	Description
No.	Score					
1	296.5	4.3	855	4	US-09-813-819-2	Sequence 2, Appl
2	296.5	4.3	855	4	US-09-920-048-2	Sequence 2, Appl
3	294.5	4.2	814	4	US-09-813-819-4	Sequence 4, Appl
4	294.5	4.2	814	4	US-09-820-048-4	Sequence 4, Appl
5	291	4.2	441	3	US-08-985-526-3	Sequence 3, Appl
6	272	3.9	656	4	US-09-351-414-2	Sequence 2, Appl
7	261	3.8	529	4	US-08-836-442-3	Sequence 2, Appl
8	261	3.8	751	2	US-08-836-443-3	Sequence 3, Appl
9	256	3.7	464	4	US-09-411-329C-14	Sequence 3, Appl
10	247	3.6	462	4	US-09-411-329C-3	Sequence 14, Appl
11	247	3.6	462	4	US-09-411-329C-17	Sequence 3, Appl
12	242	3.5	735	2	US-08-765-243-6	Sequence 17, Appl
13	242	3.5	735	5	PCR-US95-07295-6	Sequence 6, Appl
14	238.5	3.4	769	1	US-08-243-542-4	Sequence 6, Appl
15	238.5	3.4	769	1	US-08-477-407-4	Sequence 4, Appl
16	238.5	3.4	769	1	US-08-484-355-4	Sequence 4, Appl
17	235.5	3.4	1170	1	US-08-313-288B-20	Sequence 20, Appl
18	234	3.4	734	4	PCR-US95-07295-8	Sequence 8, Appl
19	234	3.4	734	5	PCR-US95-07295-8	Sequence 8, Appl
20	231	3.3	651	1	US-08-264-101-2	Sequence 2, Appl
21	231	3.3	651	2	US-08-765-243-2	Sequence 2, Appl
22	231	3.3	651	5	PCR-US95-07295-2	Sequence 2, Appl
23	229.5	3.3	391	4	US-08-706-216-6	Sequence 6, Appl
24	215	3.1	788	4	US-08-918-914-4	Sequence 4, Appl
25	214	3.1	470	1	US-08-813-150-2	Sequence 2, Appl
26	213	3.1	470	1	US-08-514-014-2	Sequence 2, Appl
27	213	3.1	470	2	US-08-833-823-2	Sequence 2, Appl

```

RESULT 1
US-09-813-819-2
; Sequence 2, Application US/09813819
; Patent No. 6294368
; GENERAL INFORMATION:
; APPLICANT: MERRUKOV, Gennady et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NOCTIFIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: US$ THEREOF
; FILE REFERENCE: CI0001177
; CURRENT APPLICATION NUMBER: US/09/813,819
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ. ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ. ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Human
; US-09-813-819-2

```

```

Db 343 DFGGVNMDHSTSLIGVASSIAHELGHSLGLDHDLPNGSC-----PCGPA--- 388
QY 436 FHMRSCKLELSRYLPSYDCLLDPPDPAMPQPELPGINYS-----MDQCRDFGSGY 490
Db 389 -----PAKTCIMEASTD-----FLPGLNFSNCSRALEKALD---GM 423
QY 491 QTCIAFRT-----FEPCKOLMCSHPDNPYFCKTKKGPPLDGTGC--APGKW 534
Db 424 GSCLEFRLPSLPRMAAFGCMNFVEPGEQDCGFLD---C---VDPCCDSLTCOLRPGAQ 477
QY 535 CFK-GHCIMKSPQTYGQDGGMSWTKFGSCS--RSCGGVRSRSGCNSPAYG-GRP 590
Db 478 CASDGPCC---QNCOLRPSGWOCRPTRGDCDLPEFCBGD---SSQC--PPDVSIGDGP 528
QY 591 CLGPMFEYOVCNSECPCGTEDFRAOOC-----AKRNSY-YVHQN 629
Db 529 CAGGQ---AVCMHGRG-ASY---AQCCOSLMGPGAOPAPPLCLOTANTRGNAFSGCGRN 580
QY 630 AKHSWVPEPDD-----DAQCELIQ--SADTGDVY 659
Db 581 PGGSYVCTPRDAICGQLCQGTGRTQPLGSIIDLLMETIDVNGTELNCMWHLDLGSDV 640
QY 660 FMOQVVDHGTGRCSTRDYVSCARGECVPYGCCKEYGSMAKDDKC---GYCGDNSHCRTV 716
Db 641 AQLPLTLPGTAGC---PGLVCIDHRCQRYDL---IGAQCRSKCHGHVC-DSNRHCYCE 693
QY 717 KG 718
Db 694 EG 695

```

## RESULT 2

```

US-09-920-048-2
; Sequence 2, Application US/09920048
; Patent No. 6344352
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001177DIY
; CURRENT APPLICATION NUMBER: US/09/920,048
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/813,819
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Human
; 09-920-048-2

```

## Query Match

```

4.3%; Score 296.5; DB 4; Length 855;
Best Local Similarity 23.7%; Pred. No. 1,1e-13;
Matches 185; Conservative 72; Mismatches 288; Indels 237; Gaps 43;

```

```

QY 53 LSHVYSGPAAASAGSMVYDPTPLRHSLSLRVARSPLHPC-----GTLWP 98
Db 35 LFMGVLPAGG-----TEEQAESEKAPREPLEPYQLDDLPISLKKVLOTSLP 83
QY 99 GRVGRISLYFNTVYCKELHLRLRNRLVYVGSSEVEM--QEDFRELROPLROCVYTG 156
Db 84 -----EPLRIKILEDDSHLELLONRELVEGPRPLVWYQPDGTVYVSEGHLEMCCTOG 138
QY 157 GVTGMGAVALAISNCDGLAGL-IRTDSTDFIE-----PLRGOQEKRSAGRTYV 206
Db 139 RVAGYAGSVNSICTGSGGLAGLVLLPERSTYLEQGGDLQGPPIISRIQDHLPLHGTAL 198
QY 207 YRREAV--QOEWAEPDGLHNEAFGLGLDPLNLGLVGQLGTEKKRRIAKGSISIVL 264

```

```

Db 199 SMRESVHTQTPPEHPLQGRH-----IRRRRDVTEETKVELV 235
QY 265 LVYDDS-VYRPHGKEHYQNYVLTLMNIYDELYIHDESIGVHINILVYLIMVYGRQSLI 323
Db 236 IYADHSEAKYRDFQHLNRLTEVALLDITFERP---LNRVALVGLAETQR---DLV 288
QY 324 E-RGNPSRSLBOYCRW--AHSQORODPSHAEDHNVFLTRQDF-GRS-GYAPYGMCHP 378
Db 289 EISNPAYLTLENFLMKRAHLLPLRP-----HDSAQLVITSTSGPYVGMAIGNSICSP 342
QY 379 LRSCALNHEDGFS--SAFVIAHETGHVLMGHEHOGNGCADEFTSLGSVAPLYOAAFHR 435
Db 343 DFGGVNMDHSTSLIGVASSIAHELGHSLGLDHDLPNGSC-----PCGPA--- 388
QY 436 FHMRSCKLELSRYLPSYDCLLDPPDPAMPQPELPGINYS-----MDQCRDFGSGY 490
Db 389 -----PAKTCIMEASTD-----FLPGLNFSNCSRALEKALD---GM 423
QY 491 QTCIAFRT-----FEPCKOLMCSHPDNPYFCKTKKGPPLDGTGC--APGKW 534
Db 424 GSCLEFRLPSLPRMAAFGCMNFVEPGEQDCGFLD---C---VDPCCDSLTCOLRPGAQ 477
QY 535 CFK-GHCIMKSPQTYGQDGGMSWTKFGSCS--RSCGGVRSRSGCNSPAYG-GRP 590
Db 478 CASDGPCC---QNCOLRPSGWOCRPTRGDCDLPEFCBGD---SSQC--PPDVSIGDGP 528
QY 591 CLGPMFEYOVCNSECPCGTEDFRAOOC-----AKRNSY-YVHQN 629
Db 529 CAGGQ---AVCMHGRG-ASY---AQCCOSLMGPGAOPAPPLCLOTANTRGNAFSGCGRN 580
QY 630 AKHSWVPEPDD-----DAQCELIQ--SADTGDVY 659
Db 581 PGGSYVCTPRDAICGQLCQGTGRTQPLGSIIDLLMETIDVNGTELNCMWHLDLGSDV 640
QY 660 FMOQVVDHGTGRCSTRDYVSCARGECVPYGCCKEYGSMAKDDKC---GYCGDNSHCRTV 716
Db 641 AQLPLTLPGTAGC---PGLVCIDHRCQRYDL---IGAQCRSKCHGHVC-DSNRHCYCE 693
QY 717 KG 718
Db 694 EG 695

```

## RESULT 3

```

US-09-813-819-4
; Sequence 4, Application US/09813819
; Patent No. 6294368
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001177
; CURRENT APPLICATION NUMBER: US/09/813,819
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 814
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(814)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-813-819-4

```

## Query Match

```

4.2%; Score 294.5; DB 4; Length 814;
Best Local Similarity 24.2%; Pred. No. 1,4e-13;
Matches 170; Conservative 67; Mismatches 258; Indels 207; Gaps 40;

```

```

QY 119 LRLRNRLVYVGSSEVEM--QEDFRELROPLROCVYTGVTGMGAVALAISNCDGLAG 176

```

```

Db      82 LELLQNLRELVPGRPLVWYQDPGRVYSEGTLENCYQGVNRYAGSWSICTGSLRG 141
QY      177 L-IRFDSIDFEIE-----PLERGOQKEASGRTHVYRRAV--QOEMAPDGLH 224
Db      142 LVLLPERSYSTLEQPGDLOGPPIISRTQDLHGHICALSWRESVHTQPPHPLGQRH 201
QY      225 NEAFGLGDLPLNLGLVGDQDLDGTERKRRHAKPGSYSTEVLVYVDS-VVRFHCKEYQNT 283
Db      202 -----IRRRDQVTEKTYELVIVADHSAQKYRDFQHLNR 238
QY      284 VLTLMNIVDELYHDESLGVHINIALVRLIMVGRSLSTIE-RGNPSRSLQYCRW-AH 340
Db      239 TLEVALLDITFEFRP-----LVNRYVALVGLAMTOR--DIVEISPNPAVLLENFLHWRRAH 291
QY      341 SQQRQDPSHAHHHDHVFELTRQDF-GPYAVTGMCHELRCALNHEDGFS--SAFY 395
Db      292 LILRLP-----HDSAQLVITGSTSPGPTVGMATONISCSDFSGGVNMDHSTSLGVASS 345
QY      396 IAHETGVLGMEHDGOGNGCADETSLSGVAPLVQAAFHFRHWSRCSKLELSRYLPSTYDC 455
Db      346 IAHETGSLGIDHDLPGNSC-----PCGPA-----PAKTC 376
QY      456 LLDPPRPAMPQPELPGINYS-----MDQCPRFEGSGYQTCIAERT-----498
Db      377 TMEASTD-----FLPGLNFSNCSRRALKEALD--GMSGCLFERLPSLPRMAAFGN 426
QY      499 --FPCQOLMCSHPDNPFCKTKGPRLDGTEC--APGKWCER-GHCIMKSPQRTYGGDC 553
Db      427 MEVEPRGQCCGFLD---C---VDPCCDSLTCQLRGAQASDGPCC---ONQLRPS 476
QY      554 GWSWTFKFGSCS--RSCGGVRSRSCNPNPSPAYG-GRPCLGPMFEYOVNSECPCPTY 610
Db      477 GWCORPRTRGCDLPERCPGD-----SSQC--PRVSLDGBPCAGGQ---AVCHMGR-ASY 527
QY      611 EDFAQOC-----AKRNSY-VYHONAKHSWVPEPDD-----641
Db      528 ----AQCQSLMGPGAAPAPLCLQTANTRGNAFGSCGRNPSGSYSCTPRDAICGQLQC 583
QY      642 -----DAQKCELICQ--SADTGADVFMNQVYVHDGTRCSYRDPYSY 679
Db      584 QTGRTPILGSIKIDLMETIDVNGTELNCWSVHLDLGSVDAQPLILTRGTACG---PGLV 640
QY      680 CARCECVPGCDEKVEGSMKADKC---GYCGDNSHCRTYKG 718
Db      641 CIDHRCQAVDL---LGAQECRSKCHGHGVC-DSNRHCYCEEG 678
BOLT 4
S-09-920-048-4
; Sequence 4, Application US/09920048
; Patent No. 6344352
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001177DIY
; CURRENT APPLICATION NUMBER: US/09/920,048
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/813,819
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 814
; TYPE: PRT
; ORGANISM: Human
; NAME/KEY: VARIANT
; LOCATION: (1)....(814)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-920-048-4

```

```

Query Match      4.2% Score 294.5; DB 4; Length 814;
Best Local Similarity 24.2%; Pred. No. 1,4e-13;
Matches 170; Conservative 67; Mismatches 258; Indels 207; Gaps 40;

QY      119 LILRPRLVWPGSSVEM--QEDFRELFPRLQRECVYTGGVGMGAVALISNCGLAG 176
Db      82 LELLQNLRELVPGRPLVWYQDPGRVYSEGTLENCYQGVNRYAGSWSICTGSLRG 141
QY      177 L-IRFDSIDFEIE-----PLERGOQKEASGRTHVYRRAV--QOEMAPDGLH 224
Db      142 LVLLPERSYSTLEQPGDLOGPPIISRTQDLHGHICALSWRESVHTQPPHPLGQRH 201
QY      225 NEAFGLGDLPLNLGLVGDQDLDGTERKRRHAKPGSYSTEVLVYVDS-VVRFHCKEYQNT 283
Db      202 -----IRRRDQVTEKTYELVIVADHSAQKYRDFQHLNR 238
QY      284 VLTLMNIVDELYHDESLGVHINIALVRLIMVGRSLSTIE-RGNPSRSLQYCRW-AH 340
Db      239 TLEVALLDITFEFRP-----LVNRYVALVGLAMTOR--DIVEISPNPAVLLENFLHWRRAH 291
QY      341 SQQRQDPSHAHHHDHVFELTRQDF-GPYAVTGMCHELRCALNHEDGFS--SAFY 395
Db      292 LILRLP-----HDSAQLVITGSTSPGPTVGMATONISCSDFSGGVNMDHSTSLGVASS 345
QY      396 IAHETGVLGMEHDGOGNGCADETSLSGVAPLVQAAFHFRHWSRCSKLELSRYLPSTYDC 455
Db      346 IAHETGSLGIDHDLPGNSC-----PCGPA-----PAKTC 376
QY      456 LLDPPRPAMPQPELPGINYS-----MDQCPRFEGSGYQTCIAERT-----498
Db      377 TMEASTD-----FLPGLNFSNCSRRALKEALD--GMSGCLFERLPSLPRMAAFGN 426
QY      499 --FPCQOLMCSHPDNPFCKTKGPRLDGTEC--APGKWCER-GHCIMKSPQRTYGGDC 553
Db      427 MEVEPRGQCCGFLD---C---VDPCCDSLTCQLRGAQASDGPCC---ONQLRPS 476
QY      554 GWSWTFKFGSCS--RSCGGVRSRSCNPNPSPAYG-GRPCLGPMFEYOVNSECPCPTY 610
Db      477 GWCORPRTRGCDLPERCPGD-----SSQC--PRVSLDGBPCAGGQ---AVCHMGR-ASY 527
QY      611 EDFAQOC-----AKRNSY-VYHONAKHSWVPEPDD-----641
Db      528 ----AQCQSLMGPGAAPAPLCLQTANTRGNAFGSCGRNPSGSYSCTPRDAICGQLQC 583
QY      642 -----DAQKCELICQ--SADTGADVFMNQVYVHDGTRCSYRDPYSY 679
Db      584 QTGRTPILGSIKIDLMETIDVNGTELNCWSVHLDLGSVDAQPLILTRGTACG---PGLV 640
QY      680 CARCECVPGCDEKVEGSMKADKC---GYCGDNSHCRTYKG 718
Db      641 CIDHRCQAVDL---LGAQECRSKCHGHGVC-DSNRHCYCEEG 678
RESULT 5
US-08-985-526-3
; Sequence 3, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Connolly, Boyle, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentia Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
08-985-526-3

```

```

Query Match          4.2%; Score 291; DB 3; Length 441;
Best Local Similarity 20.1%; Pred. No. 1.1e-13;
Matches 119; Conservative 49; Mismatches 176; Indels 248; Gaps 24;

```

```

QY 473 GINSMDEQ-----CRFDGSGYOTCLAFRTFPCQOLMCSHPDNFYCKTKKPLDGT 527
DB 22 GVQYRNNEWTDVSC-----TECHCQNSVITCKKVCSC-PI-MP-CGNATVP-DG- 66
QY 528 ECAPGKWCFCGHCIWSP----- 545
DB 67 ECCPRCPMSADADMGSPSEMTSCSTSCGNIQGRSCDSLNNRCGSSVOTRTCHIQ 126
QY 546 --EQTGODGWSWTKFSGSRSCGGVRSRSCNNPSPAYGAPCLGPMFEYOVNS 603
DB 127 ECDRFRKODGWSWMSWSSCSTYCGGVITRTICLSPSPQNGKRCBEARETACKK 186
QY 604 EECGTYEDFPRAQOCARNSYIYHONAKHSWVPEPDD-----DAQCELCIOSA-- 653
DB 187 DACP-----INGWGPWSPMDICSVTCGGGVGRSRRLCYDSRM 224
QY 654 DTGCVFENQV-----VHDSRCSYRDPYVYCARGECVPGCDKEVSGMKADKGVCG 707
DB 225 TEENKELANELRRPPLCYHNGVYRNNEWTV-----DSCTEC- 262
QY 708 GDNSHCRTVKGTLGKASKOGALXVIOIPAGARHIQLEALEKSPHSRVKNQVTFILN 767
DB 263 --HCQNSVITCKKVCSCPI-MP-CGNATVPDQ----- 289
QY 768 PKGKEATSRITVANGLEMDAVERDAKESLTSGLPFEALIALPTEGGRSSLAYKV 827
DB 290 --ECCPRCPMS-----DSADG----- 304
QY 828 IHEDLPLIGSNVLEEMDTYEMA-LKSNAPCSKACGGGIGFTKYGC--RRRRDHNV 883
DB 305 -----WSPMSEMTSCSTSCGNGIQGRSCDSLNNRCGSSV 341
QY 884 QRHLCDHKRRPKPIRRRCNQHPCSPVWV--TEEWAGCSRSGKIGVOTRGICLLPLSNG 942
DB 342 QTRCHIQ-----CDKRFKODGWSWSPSSCSTVCGD--GVITR-----ITLCNS 387
QY 943 THKVPAPACAGDRPEARPCRLVPCPAQWHLGAMSQ--CSATCGEGIGIOR 991
DB 388 PSPQNGKRCBEARET-KACKKACPLNGMGWSPMDICSVTCGGGVGR 438

```

```

RESULT 6
US-09-351-414-2
; Sequence 2, Application US/09351414
; Patent No. 6265199

```

```

; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Baidur, Nand
; APPLICANT: Delsner, Theresa A.
; APPLICANT: Bishop, Paul D.
; TITLE OF INVENTION: DISINTEGRIN HOMOLOG
; FILE REFERENCE: 98-29
; CURRENT APPLICATION NUMBER: US/09/351,414
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 696
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-351-414-2

```

```

Query Match          3.9%; Score 272; DB 4; Length 696;
Best Local Similarity 21.3%; Pred. No. 5.3e-12;
Matches 149; Conservative 102; Mismatches 247; Indels 200; Gaps 35;

```

```

QY 108 FNVTVGKELHLRLPNRLVPGSSVWMDPFLRQPLR-----QECVYTGAVTG 160
DB 30 FOIEAFSGKFLIDLILNGLI-----SSDYVEIHENKPOYSKGEHCYHSGIRG 81
QY 161 MGAVALISNDGGLAGLIRTDSTDFIEPLERGOEASGRTVAVYRREAVQOEMAPD 220
DB 82 VADSVVALSTGNGLHGMEDDTFYMIPELPLVYDEK-STGRPHLIQKTLAGVYS----- 135
QY 221 GDLINEAFGLDPLNLGLVGDQ-----LQDTERKRNHAP--GSYS-----EVLVVD 268
DB 136 -----KOMKMLTMRGDQWPFLELQMLKRRKRAVNSGIEEMKYLELMGND 185
QY 269 DSVVRFH--GREHVNQYLLTLMNIVDELYHDESLGVHINIALVRLIMVYRQSLIERG 326
DB 186 HKTYKHKHNSHAHTNFAKSVYNLVDSITYKEQ-----LNTRVLVAVETWEXQIDITT 240
QY 327 NPSRSLQGYCRMAHSGQRODPSEAHHNDHVVFLTRQDF--GPSGYAPYTGACHPRLSCAL 384
DB 241 NPVQMLHEFSKY--RQR-----IKOHADAVHLISRVTFHYKRSSLIYEGVCSRTRGVY 293
QY 385 NHEDGFSSAF--VIAHETGHVLGMHDGQ-----NGC-ADETSLSGVMAPLVQ 430
DB 294 N-EYGLPMVAQVLSQSLAQNLGLOMPESSRKPKCDCTESMGCLMETGVY----- 344
QY 431 AAFHREHMSRCSKLELSRYLP--SYDCLLDDP--FDPAMPQRPPLPGIINSMDEQCRFD 485
DB 345 --HSRKFSGKSLIEYRDLQRGCGACLENRPKLFLEPT-----EC--- 382
QY 486 FSGYOTCLAFRTFPCQKQKMSHEDNRYFCKTKKGRPLDGTGECAPGKWCFCGHCIW-- 542
DB 383 -GNGY-----VEAGEBCDCGFHYECYGLCKCKCSLSNGHAGSDGPPCCNNTSCLFPR 433
QY 543 -----KSPQYVGDGQSGSWTKFSGCSRSG--GGVRSRS 576
DB 434 GYECRDVAVNECDITTEYCTGDSQCPNHLHKQD-----YACNOQGRCYNGECKTRD 485
QY 577 RSCNNPSPAYGAPCLGPMFEYOVNSSECPGTIEDFPRAQOCARNSYIYHONAKH-- 632
DB 486 NQCOY--IWGTAKAGSDCFEYKELNTE--GTEKQ--NCGKAGDRT--QCSKHDVFC 534
QY 633 -----SWVYEPDDQOKELI-----COSA-----DTGDIVFNMQVYHDTGR 670
DB 535 GFLLCNTLRRAPRIGOLGGLIPTSTYHOGRAVYDSCGAHVAVLDDTDVGY--VEDGTP 590
QY 671 CSYRDPYVYCARGECVPGCDKEVSGMKADKGVCGG 708
DB 591 CG---PSMCLDRKCLQIOA-LNMSSCPPLDSKGVKGVSG 624

```

```

RESULT 7
US-08-836-442-3

```



```

Db      136  DRKQVHEHKYILEYLVLDNGGEKRYNENQDEIRKRVEMAYVNMV - - KKLNTH - - VAL 251
QY      309  VFLIIVAGVROSLIERENPNSRSLDQVCRMAHSSQQRQDPSPAEHHHDVFLTRDDFSPS - 367
Db      232  VGMELWTDKDKIKITP - - NASFTLENFESKMGSVL - - -SRKKHDDIAQLITATELAGTT 305
QY      368  -GVAFTVMGCHPLR -SCALINHEDG -FSSAFVIAHETGHVLCMEHDGONGCADETS -TGS 422
Db      306  VGLAMSTMGSPSYSGVGVQDDSDNLKRVAGMAHEHMGHINCFMFDYISCKCPSTICYMDK 365
QY      424  VMAPLVQAFLRPFHMSRCSKLELSRYLPS - -YDCLLDPPDPAMPORPELPGIYMSDEQ 481
Db      366  ALSEFIPTDF - - -SSCSRSLYDKFEFDKLSNCLFNAPLPTDITISPIGNQLVEMGED 420
QY      482  CRFDGSGYQICLAFRIPEEPKQJMCSPHDMPYCTCKTKKGRPLDGTCAAGKMCFKHCT 541
Db      421  C - -DGGIS -ETICKIKATFQ - - - -CALGECCECKQFKKA - - -GMCQRA - - -KDEC - 462
QY      542  WKSPQYTGQDGGWSSWTKFSGCSRSRSGGGVRSRSCNNPSPAYGGRPL - - -GPVFEY 598
Db      463  -DLPEMCG - - - - -KSGNC - - - - -PDDRQVANGFPGHGGHLMGTCPLOE 503
QY      599  QVCSNECPCTYEDPFAQCAKRSY 624
Db      504  Q-CTELWGPGRRTNPPCACAKENHF 528

RESULT      8
US-08-836-443-3
: Sequence 3, Application US/08836443
: Patent No. 5883241
: GENERAL INFORMATION:
: APPLICANT: DOCHERTY, Andrew, J.P
: APPLICANT: SLOCOMBE, Patrick, M.
: TITLE OF INVENTION: DNA SEQUENCES CODING FOR A
: TITLE OF INVENTION: HUMAN METALLOPROTEINASE AND VARIANTS THEREOF
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
: STREET: 130 Water Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/836,443
: FILING DATE: 01-MAY-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 9709420
: FILING DATE: 05-SEP-1997
: APPLICATION NUMBER: 9612145.4
: FILING DATE: 11-JUN-1996
: APPLICATION NUMBER: 9526230.9
: FILING DATE: 21-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Resnick, David S
: REGISTRATION NUMBER: 34,235
: REFERENCE/DOCKET NUMBER: 47424
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-523-3400
: TELEFAX: 617-523-6440
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 751 amino acids
: TYPE: amino acid
: STRANDEDNESS: single

```

TOPOLOGY: linear  
MOLECULE TYPE: NO. 5883241e  
US-08-836-443-3

Query Match 3.8%; Score 261; DB 2; Length 751;  
Best Local Similarity 25.1%; Pred. No. 3.8e-11;  
Matches 136; Conservative 67; Mismatches 201; Indels 136; Gaps 36;

247 TERRRRHAKPGSYS-----IEVLLVVDSDVVR-----HGKEHVONVYLLTMNIYDEIYHDE 298  
19 TKRRPRRRKEDLNSMKYVELLYVAD---YLEFQKNRRDQATGKHKLIEIANYDKERYR-- 74  
299 SLGVHINAIYRLIMVGYROSLSLIERGNSRSLEQYCRMAHSQOQDPSEAHHDHVVYR 358  
75 --SINIRIALVGLVWTHGNMCEVSE--NRYSTL---WFLSWRRKLLAQKTHDQAOL 125  
359 LTRDQGPS--GYAPVTGMCPLRSCALN---HEDGFSAFVIAHETGHVIGMEHDGOGN 413  
126 ITGMSFHTTIGLAPLMAKCSYSGGVNMDHSENAIGVATMAHEMGNFGMTHDSADC 185  
414 GCADETSLGSYMAPLVQAARHFRH--WSKSKLELSRYLPSYD--CLDDPDPAMPQPP 469  
186 CSASADGCGICMA---AATGHPFKVFNCGNRRELDRLYQSGGCMCLSNMP-DTRMLYGG 241  
470 ELPGINISMD-EQCRDPSGSGYO---TCLAFR-TEEP---CKQIMCSHP----- 510  
242 RRCGNGYDEDEEC--DGEERECNPNOCNNSCTLRPGAECANGSCCHCKLLAPGTLIC 299  
511 -----DNPFYCKTKRGP-----PLDTECAPGK-WCFKGHCITKSPQYTGDDGWW 555  
300 REQARQCDLPFEC-TGKSPHCPTFNFGMDGTPCEGQAYCYNMGCL-----TYQE----- 348  
556 SSWTKFGCSRSQGGVARSRSRSC-----NNSPAYG--GRCCIGPMNEYOVQNSEECR-G 608  
349 -----QCOQIMGEGARPADLCEKYNVAGDTFGNCKGMAN---EHRCKNRDCKG 398  
609 TYEDFRAQCCAKRNSYVYHONAKHSWVPEBDDAOKCELICQSA-----DTGDVYVFM 661  
399 KI-----QCCSEARPRLESNA---VPIDITIMNGQIQCRGTHYRGPEEGDMLDP 448  
662 NOVYHDTGRSTRDYPSYVANGEC-----VPVCGDKREVSKAKDDKC---GVCGGD-NS 711  
449 G-LVMTGTRKCGYN--HICFEGOCRNMSFFETEGCK-----KONGHYCVNNQNC 495  
712 HC 713  
496 HC 497

RESULT 9  
US-09-411-329C-14  
Sequence 14; Application US/09411329C  
Patent No. 6261820  
GENERAL INFORMATION:  
APPLICANT: Boone, Thomas  
APPLICANT: Li, Huimin  
TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE  
FILE REFERENCE: A-596  
CURRENT APPLICATION NUMBER: US/09/411.329C  
CURRENT FILING DATE: 1999-10-01  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 14  
LENGTH: 464  
TYPE: PRT  
ORGANISM: Agkistrodon contortrix  
NAME/KEY: misc.feature  
OTHER INFORMATION: Native pro-fibrinolase of Agkistrodon contortrix  
US-09-411-329C-14

Query Match 3.7%; Score 256; DB 4; Length 464;  
Best Local Similarity 24.7%; Pred. No. 4.6e-11;  
Matches 98; Conservative 67; Mismatches 158; Indels 74; Gaps 19;

92 PGGLMPGRVGRHSILYFVNTYFGKELHLRLRPNRLVYVG--SSVEMQEDFRELFRQPLRQ 150  
114 PRGAVQPKY--EDAMQYEFKVNSEPVVLLHKKKGLFSEDSYETHSPDGREITTYPLGE 171  
151 E-CVYTGCVTGMPGAANAISNCDGLAFLRTDSIDFIEPLERQOEKAEAGSRHHVYRR 209  
172 DHCYHGRLENDADASTASISACNGLKGFKLQGEWYLLIEPLSDSE-----AAHAYKY 225  
210 EAVOQEMAEPPDGLDHNFAFGLDLPNLGLVGD-----QLGDTERRRRHAKPGS 258  
226 ENVEKE-----DEAPKMGVGNMSEYEPKAPQNLNTRKQGRF--PQR 268  
259 YSIEVLLVVDSDV-VREHG-KEHVONVYLLTMNIYDEIYHDESLGCHINIALVRLIMGY 316  
269 Y-VQIVIVADHHRMNTKYNQSDSKIRQWVHQLVNTINEIYRP---LNTQFLVGLTISN 323  
317 ROSLSLIERGNSRSLEQYCRMAHSQOQDPSEAHHDHVVYFLTRDQGPS--GYAPYTG 374  
324 QDLITVTSYSHD--TLASFGNM---RETDLLRQRHDAQLLTAIDPDGDTGLAYYGG 377  
375 MCHPLRSCALNHEDGFSS-----AFVIAHETGHVIGMEHDGOGNCADETSLGSYV 425  
378 MCQ-----LKHSTGVIOHSAINLLVALLTMAHELGHNMGNDGNOCGANS---CYM 428  
426 APLVQAARHFRHRSRCKLELSRYLP--SYDCLDDP 460  
429 AAMLSDQPSKL-FSDCKRDYQTFILVNNPCLINKP 464

RESULT 10  
US-09-411-329C-3  
Sequence 3; Application US/09411329C  
Patent No. 6261820  
GENERAL INFORMATION:  
APPLICANT: Boone, Thomas  
APPLICANT: Li, Huimin  
TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE  
FILE REFERENCE: A-596  
CURRENT APPLICATION NUMBER: US/09/411.329C  
CURRENT FILING DATE: 1999-10-01  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 3  
LENGTH: 462  
TYPE: PRT  
ORGANISM: Artificial  
FEATURE:  
OTHER INFORMATION: Native pro-NAT (analog of fibrinolase)  
US-09-411-329C-3

Query Match 3.6%; Score 247; DB 4; Length 462;  
Best Local Similarity 24.4%; Pred. No. 2.1e-10;  
Matches 97; Conservative 66; Mismatches 158; Indels 76; Gaps 19;

92 PGGLMPGRVGRHSILYFVNTYFGKELHLRLRPNRLVYVG--SSVEMQEDFRELFRQPLRQ 150  
114 PRGAVQPKY--EDAMQYEFKVNSEPVVLLHKKKGLFSEDSYETHSPDGREITTYPLGE 171  
151 E-CVYTGCVTGMPGAANAISNCDGLAFLRTDSIDFIEPLERQOEKAEAGSRHHVYRR 209  
172 DHCYHGRLENDADASTASISACNGLKGFKLQGEWYLLIEPLSDSE-----AAHAYKY 225  
210 EAVOQEMAEPPDGLDHNFAFGLDLPNLGLVGD-----QLGDTERRRRHAKPGS 258  
226 ENVEKE-----DEAPKMGVGNMSEYEPKAPQNLNTRKQGRF--PQR 266

Thu Aug 29 10:45:43 2002

us-09-938-330-20.raii

Page 7

```

QY 259 YSFEVLVVDGVS-VRFNG-KEHONVVLMLIVDELIYHDSLSGVHINILVALLVWG 316
D 267 Y-VQVIVADHMAKNTKIGSDSKIROWHODVYVITINELTR---LNIQFTVLEITMSN 321
QY 317 ROSLSIERGNESSRLSEOVCSMAHSOORODPSHAENHNAVFLTRKDFGSPS--GYAPVYG 374
D 322 ODILITVSVSHD--PLASFGNW---RETDLERQGHDAQGLITLIDPDSGTVGLAVYGG 375
QY 375 MCHPLRSCALNHDGFS-----AAYIAHEFGVILGMEHDSQNGCADETSLGYSV 425
D 376 MCO-----LKSITGVITODHSAILNLVALVITMAHEHLNGLNMGNDGNCGANS--CVV 426
QY 426 APVQAAHFHFHWSRCSKLTSLRYLP--SYVDLDDP 460
D 427 AAMLSDPSKSL-FSDCSKXKDYQFFLYVNNPQCLINKP 462

```

Arch completed: August 28, 2002, 20:11:08  
Job time: 865 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: August 28, 2002, 23:53:34 ; Search time 32.43 seconds  
(without alignments)  
158.168 Million cell updates/sec

Title: US-09-938-330-22-COPY\_1040\_1249  
Perfect score: 1206  
Sequence: 1 PEGQWPGSGPLHPIINKISS.....PTGPTLCVLPDLSQRLGHT 210

Indexing table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: /usr2/6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/PCRTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	175	14.5	761	2	US-08-707-237A-84
2	175	14.5	762	1	US-08-642-255-114
3	175	14.5	762	1	US-08-642-255-120
4	175	14.5	762	1	US-08-397-633A-26
5	175	14.5	762	1	US-08-397-633A-31
6	158	13.1	682	1	US-08-642-255-126
7	158	13.1	682	1	US-08-397-633A-36
8	154	12.8	417	1	US-08-175-155-69
9	154	12.8	417	1	US-08-477-509B-104
10	154	12.8	417	1	US-08-642-255-102
11	154	12.8	417	2	US-08-707-237A-76
12	154	12.8	417	3	US-08-482-085B-104
13	154	12.8	417	4	US-08-444-791A-104
14	154	12.8	837	1	US-08-175-155-68
15	154	12.8	837	1	US-08-477-509B-103
16	154	12.8	837	1	US-08-642-255-101
17	154	12.8	837	2	US-08-707-237A-75
18	154	12.8	837	2	US-08-482-085B-103
19	154	12.8	837	4	US-09-444-791A-103
20	154	12.8	897	1	US-08-397-633A-50
21	152.5	12.6	829	1	US-08-642-255-132
22	152.5	12.6	829	4	US-09-219-849-5
23	151.5	12.6	960	4	US-08-397-633A-53
24	150	12.4	330	1	US-08-642-255-32
25	150	12.4	408	1	US-07-609-716-65
26	150	12.4	408	4	US-08-475-411A-65
27	150	12.4	408	4	US-08-478-029A-65

28	145.5	12.1	1255	2	US-09-080-897-4	Sequence 4, Appl
29	145.5	12.1	1255	3	US-08-899-595-1	Sequence 1, Appl
30	145.5	12.1	1255	4	US-09-323-735-4	Sequence 4, Appl
31	145	12.0	623	4	US-09-029-348-3	Sequence 3, Appl
32	145	12.0	626	4	US-09-029-348-2	Sequence 2, Appl
33	143.5	11.9	357	1	US-07-609-716-66	Sequence 66, Appl
34	143.5	11.9	357	1	US-08-642-255-33	Sequence 33, Appl
35	143.5	11.9	357	4	US-08-475-411A-66	Sequence 66, Appl
36	143.5	11.9	357	4	US-08-478-029A-66	Sequence 66, Appl
37	137	11.4	1248	2	US-09-080-897-2	Sequence 2, Appl
38	137	11.4	1248	4	US-09-323-735-2	Sequence 2, Appl
39	137	11.4	1341	3	US-08-963-825-18	Sequence 18, Appl
40	137	11.4	1341	4	US-09-570-573-18	Sequence 18, Appl
41	137	11.4	1341	4	US-09-548-608-18	Sequence 18, Appl
42	136	11.3	1442	2	US-08-316-650-12	Sequence 12, Appl
43	136	11.3	1442	5	PCT-US95-02251-12	Sequence 12, Appl
44	135	11.2	1274	4	US-09-095-443-2	Sequence 2, Appl
45	134.5	11.2	1315	3	US-08-899-595-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-08-707-237A-84  
: Sequence 84, Application US/08707237A  
: Patent No. 5830713  
: GENERAL INFORMATION:  
: APPLICANT: Ferrari, Franco A.  
: APPLICANT: Capello, Joseph  
: APPLICANT: Crisman, John W.  
: APPLICANT: Dorman, Mary A.  
: TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC  
: NUMBER OF SEQUENCES: 108  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
: STREET: Four Embarcadero Center, Suite 3400  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: United States  
: ZIP: 94111-4187  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/707,237A  
: FILING DATE: 03-SEP-1996  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/175,155  
: FILING DATE: 29-DEC-1993  
: APPLICATION NUMBER: US 08/053,049  
: FILING DATE: 22-APR-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/609,716  
: FILING DATE: 06-NOV-1990  
: APPLICATION NUMBER: US 07/269,429  
: FILING DATE: 09-NOV-1988  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/114,618  
: FILING DATE: 29-OCT-1987  
: APPLICATION DATA:  
: APPLICATION NUMBER: US 06/927,258  
: FILING DATE: 04-NOV-1986  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Treccartin, Richard F.  
: REGISTRATION NUMBER: 31,801  
: REFERENCE/DOCKET NUMBER: A-55186-10/MHD



Db 87 GOOHHLGARGO---AGDVSGPG-----APGTPGQGLPGSPG--APGTPGQGLPGSPGA 136  
 QY 153 TOTPTP-----VPEDKGGPGEGLRH-----PGTSLPADLPGRPPEPCHP 191  
 Db 137 PGTTPGQGLPGSPGAPGTTPGEGOOHHLGARGAGDVSGPAGPAGTTPGQGLPGSPGAPGTP 196

RESULT 4  
 US-08-397-633A-26  
 ; Sequence 26, Application US/08397633A  
 ; Patent No. 5773577  
 ; GENERAL INFORMATION:

APPLICANT: Cappello, Joseph  
 TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATE-ESCAPABLE  
 NUMBER OF SEQUENCES: 105  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
 STREET: 4 Embarcadero Center, Suite 3400  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111-4187

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/397,633A  
 FILING DATE:  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Rowland, Bertlam I  
 REGISTRATION NUMBER: 20,015  
 REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 781-1989  
 TELEFAX: (415) 398-3249  
 TELETYPE: 910 277299  
 INFORMATION FOR SEQ ID NO: 26:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 762 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 8-397-633A-26

Query Match 14.5%; Score 175; DB 1; Length 762;  
 Best Local Similarity 35.0%; Pred. No. 2.2e-06;  
 Matches 63; Conservative 8; Mismatches 51; Indels 58; Gaps 12;

QY 58 ASGP-----NPGPD-----PGTSLPFPSTPGSP--LPGRQDPAADAEPPGKPTG 100  
 Db 29 ASDPMGAPGTTPGQGLPGSPGAPGTTPGQGLP--GSPGAPGTTPGQGLPGSPGAPGTTPG 86  
 QY 101 SEDHONGRATQRLGALDTSPTQHPFAPETP-----IPGASWSISPTTPG--GLPW--GW 152  
 Db 87 GOOHHLGARGO---AGDVSGPG-----APGTPGQGLPGSPG--APGTPGQGLPGSPGA 136  
 QY 153 TOTPTP-----VPEDKGGPGEGLRH-----PGTSLPADLPGRPPEPCHP 191  
 Db 137 PGTTPGQGLPGSPGAPGTTPGEGOOHHLGARGAGDVSGPAGPAGTTPGQGLPGSPGAPGTP 196

RESULT 5  
 US-08-397-633A-31  
 ; Sequence 31, Application US/08397633A  
 ; Patent No. 5773577  
 ; GENERAL INFORMATION:  
 APPLICANT: Cappello, Joseph

TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATE-ESCAPABLE  
 TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING  
 NUMBER OF SEQUENCES: 105  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
 STREET: 4 Embarcadero Center, Suite 3400  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111-4187

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/397,633A  
 FILING DATE:  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Rowland, Bertlam I  
 REGISTRATION NUMBER: 20,015  
 REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 781-1989  
 TELEFAX: (415) 398-3249  
 TELETYPE: 910 277299  
 INFORMATION FOR SEQ ID NO: 31:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 762 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-397-633A-31

Query Match 14.5%; Score 175; DB 1; Length 762;  
 Best Local Similarity 35.0%; Pred. No. 2.2e-06;  
 Matches 63; Conservative 8; Mismatches 51; Indels 58; Gaps 12;

QY 58 ASGP-----NPGPD-----PGTSLPFPSTPGSP--LPGRQDPAADAEPPGKPTG 100  
 Db 29 ASDPMGAPGTTPGQGLPGSPGAPGTTPGQGLP--GSPGAPGTTPGQGLPGSPGAPGTTPG 86  
 QY 101 SEDHONGRATQRLGALDTSPTQHPFAPETP-----IPGASWSISPTTPG--GLPW--GW 152  
 Db 87 GOOHHLGARGO---AGDVSGPG-----APGTPGQGLPGSPG--APGTPGQGLPGSPGA 136  
 QY 153 TOTPTP-----VPEDKGGPGEGLRH-----PGTSLPADLPGRPPEPCHP 191  
 Db 137 PGTTPGQGLPGSPGAPGTTPGEGOOHHLGARGAGDVSGPAGPAGTTPGQGLPGSPGAPGTP 196

RESULT 6  
 US-08-642-255-126  
 ; Sequence 126, Application US/08642255  
 ; Patent No. 5773249  
 ; GENERAL INFORMATION:  
 APPLICANT: Cappello, Joseph  
 TITLE OF INVENTION: High Molecular Weight Collagen-Like  
 TITLE OF INVENTION: Protein Polymers  
 NUMBER OF SEQUENCES: 135  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
 STREET: 4 Embarcadero Center, Suite 3400  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94111-4187  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/642,255  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: ROWLAND, Bertam I.  
REGISTRATION NUMBER: 20,015  
REFERENCE/DOCKET NUMBER: A55556-3/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299 FHT UR  
INFORMATION FOR SEQ ID NO: 126:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 682 amino acids  
TYPE: amino acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
8-642-255-126

Query Match 13.1%; Score 158; DB 1; Length 682;  
Best Local Similarity 37.6%; Pred. No. 4.3e-05;  
Matches 62; Conservative 9; Mismatches 50; Indels 44; Gaps 14;

QY 58 ASGP-----NPGPD-----PGTSLPPESTPGSP-LPGOPDPAADAEPPGKPTG 100  
DB 29 ASDPMGAPGTPOGGLPGSPGAPGTTPGQGLP--GSPGAPGTGPGGLPGSPAPGTPGG 86  
QY 101 SEDHGHGRATQGLPGLDTSPTGTQ---HPAPETP---IPGASWSISPTTPG--GLPW 150  
DB 87 AK--QAGDVGS--PGAPGT--PGQGLPGSPGAPGTTPGQGLPGSPG--AGTTPGQGLPG 139  
QY 151 --GWTOTPTPYPE--DKGPGEDLRHPTSLPADLGRPEPCHP 191  
DB 140 SPGAPGTGKAKQAGDVGSPGA---PGTTPGQGLPGSPGAPGTTP 180

RESULT 7  
US-08-397-633A-36  
Sequence 36, Application US/08397633A  
Patent No. 577357  
GENERAL INFORMATION:  
APPLICANT: Cappello, Joseph  
TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATE SCAPABLE  
TITLE OF INVENTION: OF ENZYMA TIC CROSS-LINKING  
NUMBER OF SEQUENCES: 105  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/397,633A  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland, Bertam I.  
REGISTRATION NUMBER: 20,015  
REFERENCE/DOCKET NUMBER: A-58648-1/BIR PROP-011-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 682 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-397-633A-36

Query Match 13.1%; Score 158; DB 1; Length 682;  
Best Local Similarity 37.6%; Pred. No. 4.3e-05;  
Matches 62; Conservative 9; Mismatches 50; Indels 44; Gaps 14;

QY 58 ASGP-----NPGPD-----PGTSLPPESTPGSP-LPGOPDPAADAEPPGKPTG 100  
DB 29 ASDPMGAPGTPOGGLPGSPGAPGTTPGQGLP--GSPGAPGTGPGGLPGSPAPGTPGG 86  
QY 101 SEDHGHGRATQGLPGLDTSPTGTQ---HPAPETP---IPGASWSISPTTPG--GLPW 150  
DB 87 AK--QAGDVGS--PGAPGT--PGQGLPGSPGAPGTTPGQGLPGSPG--AGTTPGQGLPG 139  
QY 151 --GWTOTPTPYPE--DKGPGEDLRHPTSLPADLGRPEPCHP 191  
DB 140 SPGAPGTGKAKQAGDVGSPGA---PGTTPGQGLPGSPGAPGTTP 180

RESULT 8  
US-08-175-155-69  
Sequence 69, Application US/08175155  
Patent No. 5641648  
GENERAL INFORMATION:  
APPLICANT: Ferrari, Franco A.  
APPLICANT: Cappello, Joseph  
APPLICANT: Crissman, John W.  
APPLICANT: Dorman, Mary A.  
TITLE OF INVENTION: Methods for Preparing Synthetic  
TITLE OF INVENTION: Repetitive DNA  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/175,155  
FILING DATE: 29-DEC-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland, Bertam I.  
REGISTRATION NUMBER: 20015  
REFERENCE/DOCKET NUMBER: A-55186-5/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 417 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-175-155-69







Query Match	12.88;	Score 154;	DB 1;	Length 837;
Best Local Similarity	35.98;	Pred. No. 0.00011;		
Matches 60;	Conservative 7;	Mismatches 46;	Indels 54;	Gaps 13;
58 ASGP-----NPPGD-----PPTSLPPFTTPGSP-LPGQDDADAEPGGR-				98

Db 29 ASDPMGAPGTGGPGLPGSPGAPGTGGPGLPGSPGAPGTGGP 86  
QY 99 ---TSEEDHOGHGRATQLPALDTSSTPGTQ---HPFAPETP---IPGASMSISPTTGG- 146  
Db 87 QGLPGS-----PGAPGT--PGPQGLPGSPGAPGTGGPGLPGSPG-APGTGGP 131  
QY 147 -GLPMGWOTPTPYVED-KGQGEDLRHFGTSLPADLPGRPEPCHP 191  
Db 132 QGLPGSPGAPGTGGPGLPGSPGAPGTGGPGLPGSPGAPGTGGP 174

## RESULT 15

US-08-477-509B-103  
Sequence 103, Application US/08477509B  
Patent No. 5770697

## GENERAL INFORMATION:

APPLICANT: Ferrari, Franco A  
APPLICANT: Cappello, Joseph  
APPLICANT: Crisman, John W  
APPLICANT: Dorman, Mary A  
TITLE OF INVENTION: No. 5770697el Peptides Comprising Repetitive  
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same  
NUMBER OF SEQUENCES: 112  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94111

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,509B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/115,155  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/053,049  
FILING DATE: 22-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/114,618  
FILING DATE: 29-OCT-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/927,258  
FILING DATE: 04-NOV-1986  
ATTORNEY/AGENT INFORMATION:  
NAME: Treccartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEO ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-477-509B-103

Query Match 12.8%; Score 154; DB 1; Length 837;  
Best Local Similarity 35.9%; Pred. No. 0.00011;  
Matches 60; Conservative 7; Mismatches 46; Indels 54; Gaps 13;

QY 58 ASGP-----NPGPD-----PGPTSLPPTSTPGSP-LPGQDPPADAAEPCKP-- 98  
Db 29 ASDPMGAPGTGGPGLPGSPGAPGTGGPGLPGSPGAPGTGGP 86  
QY 99 ---TSEEDHOGHGRATQLPALDTSSTPGTQ---HPFAPETP---IPGASMSISPTTGG- 146  
Db 87 QGLPGS-----PGAPGT--PGPQGLPGSPGAPGTGGPGLPGSPG-APGTGGP 131  
QY 147 -GLPMGWOTPTPYVED-KGQGEDLRHFGTSLPADLPGRPEPCHP 191  
Db 132 QGLPGSPGAPGTGGPGLPGSPGAPGTGGPGLPGSPGAPGTGGP 174

## RESULT 16

US-08-642-255-101  
Sequence 101 Application US/08642255  
Patent No. 5773249

## GENERAL INFORMATION:

APPLICANT: CAPPELLO, Joseph  
APPLICANT: FERRARI, Franco A.  
TITLE OF INVENTION: High Molecular Weight Collagen-Like  
TITLE OF INVENTION: Protein Polymers  
NUMBER OF SEQUENCES: 135  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/642,255  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: ROWLAND, Berttram I.  
REGISTRATION NUMBER: 20,015  
REFERENCE/DOCKET NUMBER: A55556-3/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8771  
TELEFAX: (415) 494-8770  
TELEX: 910 277299 FHT UR  
INFORMATION FOR SEO ID NO: 101:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-642-255-101

Query Match 12.8%; Score 154; DB 1; Length 837;  
Best Local Similarity 35.9%; Pred. No. 0.00011;  
Matches 60; Conservative 7; Mismatches 46; Indels 54; Gaps 13;

QY 58 ASGP-----NPGPD-----PGPTSLPPTSTPGSP-LPGQDPPADAAEPCKP-- 98  
Db 29 ASDPMGAPGTGGPGLPGSPGAPGTGGPGLPGSPGAPGTGGP 86  
QY 99 ---TSEEDHOGHGRATQLPALDTSSTPGTQ---HPFAPETP---IPGASMSISPTTGG- 146  
Db 87 QGLPGS-----PGAPGT--PGPQGLPGSPGAPGTGGPGLPGSPG-APGTGGP 131  
QY 147 -GLPMGWOTPTPYVED-KGQGEDLRHFGTSLPADLPGRPEPCHP 191  
Db 132 QGLPGSPGAPGTGGPGLPGSPGAPGTGGPGLPGSPGAPGTGGP 174

RESULT 17  
US-08-707-237A-75  
Sequence 75, Application US/08707237A  
Patent No. 5830713  
GENERAL INFORMATION:  
APPLICANT: Ferrari, Franco A.  
APPLICANT: Capello, Joseph  
APPLICANT: Crisman, John W.  
APPLICANT: Dorman, Mary A.  
TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC  
TITLE OF INVENTION: REPETITIVE DNA  
NUMBER OF SEQUENCES: 108  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fleht, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/707,237A  
FILING DATE: 03-SEP-1996  
CLASSIFICATION: 435  
APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,155  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/053,049  
FILING DATE: 22-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/609,716  
FILING DATE: 06-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/269,429  
FILING DATE: 09-NOV-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/114,618  
FILING DATE: 29-OCT-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/927,258  
FILING DATE: 04-NOV-1986  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-55186-10/WHD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 75:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-707-237A-75

Query Match 12.8%, Score 154; DB 2; Length 837;  
Best Local Similarity 35.9%; Pred. No. 0.00011;  
Matches 60; Conservative 7; Mismatches 46; Indels 54; Gaps 13;

QY 58 ASGP-----NGPD-----PGTSLPPSTPSP-LPGPDADAAEPCKP-- 98  
DB 29 ASDPGACGTGPGQGLPGSPGAPGTGPGQGLP--GSPGACGTGPGQGLPGSPGAPGTGPGP 86

QY 99 ---TGSEFDHOGHATGALDITSSPQTQ----HPPAPETP-----IPGASMSISPTTGG- 146  
DB 87 QGLPGS-----PGAGT--PGQGLPGSPGAPGTGPGQGLPGSPG--ABGTGPG 131

QY 147 -GLPWGWTGTTPVED-KGQGEDLRHPTSLPADLPGRPPPCHP 191  
DB 132 QGLPGSPGACGTGPGQGLPGSPG----PBTGPGQGLPGSPGAPGT 174

RESULT 18  
US-08-482-085B-103  
Sequence 103, Application US/08482085B  
Patent No. 6018030  
GENERAL INFORMATION:  
APPLICANT: Ferrari, Franco A.  
APPLICANT: Richardson, Charles  
APPLICANT: Chambers, James  
APPLICANT: Causey, Stuart  
APPLICANT: Pollock, Thomas J.  
APPLICANT: Crisman, John W.  
TITLE OF INVENTION: No. 6018030e1 Peptides Comprising Repetitive  
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same  
NUMBER OF SEQUENCES: 112  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fleht, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,085B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
APPLICATION DATA:  
APPLICATION NUMBER: US 06/927,258  
FILING DATE: 04-NOV-1986  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/114,618  
FILING DATE: 29-OCT-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/053,049  
FILING DATE: 22-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,155  
FILING DATE: 29-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-55186-6/RFT/MTX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-482-085B-103

Query Match 12.8%, Score 154; DB 3; Length 837;  
Best Local Similarity 35.9%; Pred. No. 0.00011;  
Matches 60; Conservative 7; Mismatches 46; Indels 54; Gaps 13;



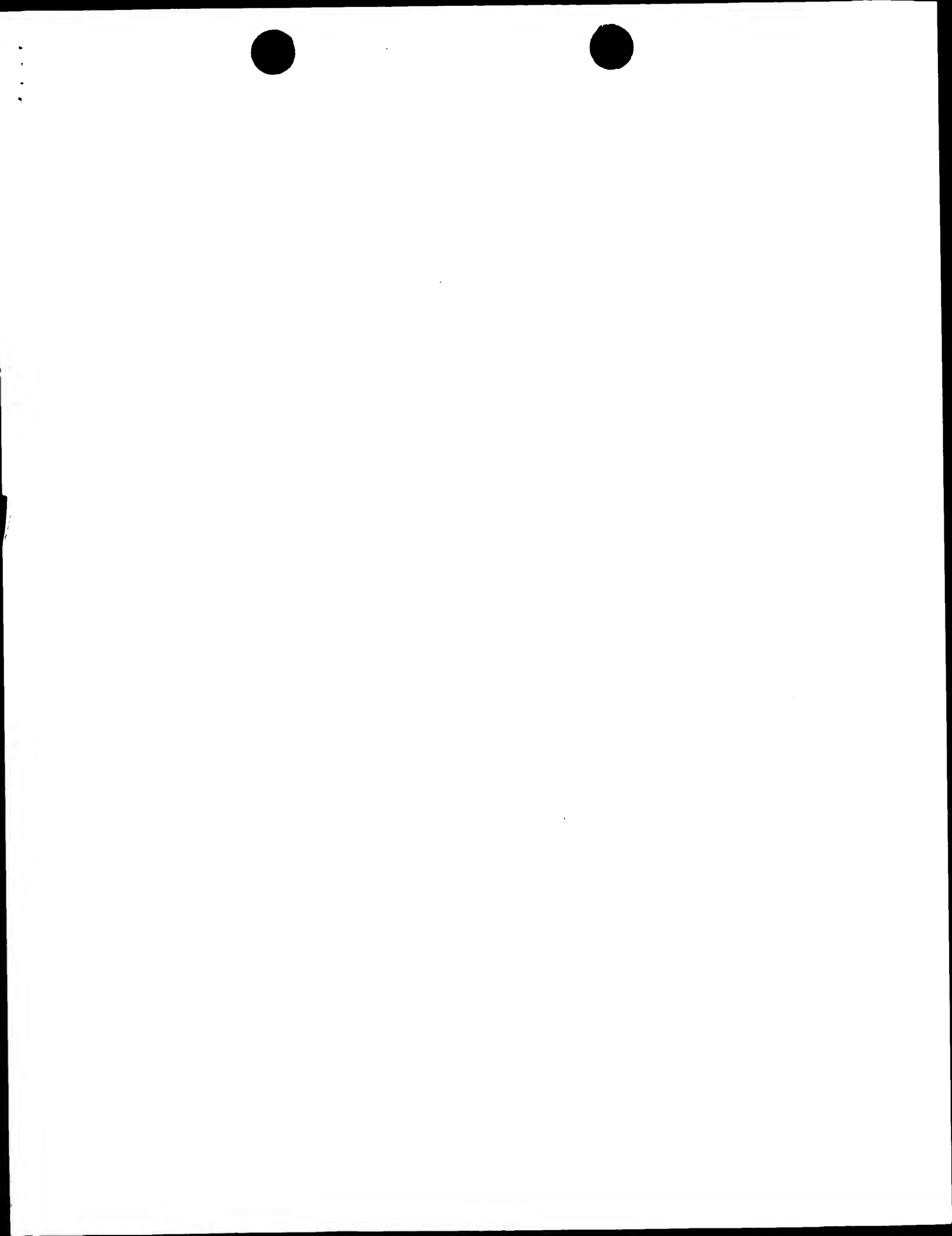
Thu Aug 29 10:46:18 2002

us-09-938-330-22\_copy\_1040\_1249.ra

Page 11

Db 132 0G1PGSPGAPGTPGPGQGLPGSPGA---PGTPGPGQGLPGSPGAPGTP 174

Search completed: August 29, 2002, 00:08:25  
Job time: 891 sec





Sequence	Strid	Org	ZScore	Escore	Len	Documentation
gb AF158666	+	6723.50	3300.95	1.7e-115	3672	AF158666 Homo sapiens a dis
gb pat:AX319856	+	6608.00	3199.91	1.2e-169	3561	AX319856 Sequence 20 from Pat
gb pat:AX342633	+	6468.50	3176.49	1.4e-168	3567	AX342633 Sequence 31 from Pat
gb pat:BX6611	+	5160.50	2537.96	3.4e-133	3144	AX211284 Sequence 1 from Pat
gb pat:BX6611	+	3625.50	1785.99	1.4e-91	3618	ES8661 Novel metalloprote
gb pat:AB002364	+	3310.50	1640.78	1.0e-90	5774	AB002364 Human mRNA for KIAA
gb com:B BPC13125	+	3399.50	1627.34	5.0e-83	4580	X96389 B. taurus mRNA for pro
gb pat:AX128433	+	1438.00	715.27	1.8e-31	5061	ES50215 Homo sapiens mRNA fo
gb pat:AX139851	+	1436.00	710.82	2.1e-31	5353	AX128433 Sequence 2 from Pat
gb pat:AX139851	+	1341.50	670.88	3.3e-29	3312	AX139851 Sequence 15 from Pat
gb pat:ES5273	+	1334.00	667.04	8.6e-29	3400	ES5273 Novel metalloprotease
gb pat:AX319852	+	1324.00	661.64	1.7e-28	3675	AF16762 Homo sapiens zinc me
gb pat:AX327752	+	1319.50	660.09	2.1e-28	3132	AX319852 Sequence 16 from Pat
gb pat:AX327747	+	1302.50	652.02	5.9e-28	3207	AX327752 Sequence 10 from Pat
gb pat:AX327745	+	1302.50	651.64	6.2e-28	3402	AX327747 Sequence 3 from Pat
gb pat:AX338537	+	1302.50	651.64	6.2e-28	3403	AX338537 Sequence 3 from Pat
gb pat:AX327755	+	1302.00	651.26	6.2e-28	3471	AX327745 Sequence 1 from Pat
gb pat:AX319857	+	1278.50	630.37	4.4e-27	5808	AX319857 Sequence 11 from Pat
gb pat:AX319857	+	1244.50	626.40	4.7e-26	5115	AX319857 Sequence 21 from Pat
gb pat:AX3261918	+	1234.50	617.90	4.7e-26	3674	AF261918 Homo sapiens dist
gb pat:AB037733	+	1232.00	614.47	7.3e-26	5139	AB037733 Homo sapiens mRNA fo
gb pat:AX319858	+	1228.50	613.61	8.2e-26	4518	AX319858 Sequence 22 from Pat
gb pat:AX319854	+	1155.00	580.80	5.2e-24	2805	AX319854 Sequence 18 from Pat
gb pat:AX319854	+	1145.00	569.85	2.2e-23	4878	AF149118 Rattus norvegicus a
gb pat:AX319854	+	1140.00	566.95	3.9e-23	4180	D67076 Homo sapiens mRNA fo
gb pat:AX342635	+	1124.50	565.99	3.9e-23	2937	AX319856 Sequence 24 from Pat
gb pat:AX342635	+	1118.00	562.43	5.8e-23	2937	AX342635 Sequence 32 from Pat
gb pat:AF170084	+	1109.00	551.88	2.2e-22	4447	AF304446 Rattus norvegicus ADP
gb pat:ES29406	+	1100.00	551.57	2.0e-22	4447	AF170084 Homo sapiens metallo
gb pat:AF2037767	+	1100.00	551.11	2.3e-22	4014	AF060152 Homo sapiens METH
gb pat:ES29406	+	1100.00	550.59	2.5e-22	4309	ES29406 Novel integrin ligand
gb pat:AF207664	+	1098.50	553.07	1.9e-22	2853	AF207664 Homo sapiens mRNA fo
gb pat:ES58656	+	1098.50	525.96	7.0e-21	172421	ES58656 Novel metalloproteas
gb pat:AC016043	+	1097.50	525.04	7.0e-21	151161	AC016043 Homo sapiens chrom

seq_name	gb_pr	AF358666																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			</
----------	-------	----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----



```

467 nProProGluLeuProGlyIleasnTyrSerMetAspGluGlnCysArgp 484
|||||
1401 GCGCCAGAGCTCCTGGGATGACACTACATGATGACAGCTGCGCT 1450
484 heasPheGlySerGlyTyrGlnThrCysLeuAlaPheArgThrPheGlu 500
|||||
1451 TTGACTTTGGCAGTGGCTACCAAGACCTGCTTGCCATTCAGAGCTTTAG 1500
501 ProCysLysGlnLeuThrPcySerSerHisProAspAsnProTyrPheCysLy 517
|||||
1501 CCTGGACAGCAGCTGTGGTGCAGCCATCTGTACAACCCGTACTTCTGCAA 1550
517 sThrLysLysGlyProProLeuAspGlyThrGluCysAlaProGlyLysT 534
1551 GACCAAGAGAGGGGCCCCGCTGGATGGAGCTAGTGTCCACCCGGCAAGT 1600
534 rPcySerPheLysGlyHisCysLleTrrPlySerProGluGlnThrTyrGly 550
1601 GGTGGTTCMAAGGTCATGTCATCTGGAAGTCGCCGAGCAGACATATGCG 1650
551 GlnAspGlyGlyTrrPserSerTrpThrLysPheGlySerCysSerArgSe 567
1651 CAGGATGAGAGGCTGAGCTCCTGGACCAAGTTGGTCTATGTTCGGGTC 1700
567 rCysGlyGlyGlyValArgSerArgSerArgSerCysAsnAsnProSerP 584
1701 ATGTGGGGGGGGGTGGCATCCCGCAGCGGAGGTGCACAAACCCCTCC 1750
584 roAlaTyrGlyGlyArgProCysLeuGlyProMetPheGluTyrGlnVal 600
1751 CAGCTATGAGAGCGCCCGTGTAGGGCCCATGTTCCAGTACCAAGTC 1800
601 CysAsnSerGluGluCysProGlyTyrThrTyrGlnAspPheArgAlaGln 617
1801 TCACAACAGGAGAGTGGCTCGGACCTACAGACACTTCGGGGCCCAAGA 1850
617 nCysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHisSer 634
1851 GGTGTCCCAAGGCAACTGTACTATGTGCACCAAGATGCCAAGCACAGCT 1900
634 rPValProTyrGluProAspAspAlaGlnLysCysGluLeuLleCys 650
1901 GGGTCCCTACGAGCTGACGATACGACCCAGAAAGTGTGAGCTGATCTGC 1950
651 GlnSerAlaAspTrpGlyAspValValPheMetAsnGlnValValHisAs 667
1951 CAGTCGGCGGACACGGGGACGTGTGTCATGAAACCAAGTGGTTCACGA 2000
667 pGlyThrArgCysSerTyrArgAspProTyrSerValCysAlaArgGlyS 684
2001 TGGACACGCTGCAGCTACCGGAGCCATACAGCGTCTGTGCCGTGGCG 2050
684 lCysValProValGlyCysAspLysGluValGlySerMetLysAlaAsp 700
2051 AGTGTGTGCTGCTGGCTGTGACAAAGAGTGGGTTCCATGAAGCGCGAT 2100
701 AspLysCysGlyValLysGlyGlyAspAsnSerHisCysArgThrValLy 717
2101 GACAAAGTGTGAGTCTCGGGGTGACCACTCCACTGCGAGGACTGTGAA 2150
717 sGlyThrLeuGlyLysAlaSerLysGlnAlaGlyAlaLeuLysLeuValG 734
2151 GGGAGCGCTGGGCAAGGCTCCCAAGCAGGAGAGAGCTTCMAAGCTGTGCT 2200
734 lInLeuProAlaGlyAlaArgHisTleGlnLleGlnAlaLeuGlnLysSer 750
2201 AGATCCAGAGAGTGCAGGACATCCAGATTGAGGACATGGAGAGTGC 2250
751 ProHisArgSerValValLysAsnGlnValThrGlySerPheLleLeuAs 767
2251 CCCACCGCATTTGTGTGAAGAACAGGTACCGGCACTTCATCTCTCAA 2300
767 nProLysGlyLysGluAlaThrSerArgThrPheThrAlaMetGlyLeuG 784
|||||
2301 CCGCAAGGGCAGGAAGCCACAAAGCGGACCTTCACCCCATAGGGCTCG 2350
784 lInrPGLuAspAlaValGluAspAlaLysGlnSerLeuLysThrSerLy 800
2351 AGTGGAGAGATGGGTGGAGATGGCCAAAGAAAGCTCAACACCAAGCGGG 2400
801 ProLeuProGluAlaLleAlaLleLeuAlaLeuProProThrGlnGlyL 817
2401 CCTGCTGCTGAAGCCATTGCCATCTGGCTCTCCCGCCCACTGAGGGTGG 2450
817 yProArgSerSerLeuAlaTyrLysTyrValLleHisGluAspLeuLeuP 834
2451 CCGCCGACAGCAGCTGGGCTTCACATGATCATCATGAGGACCTGCTGC 2500
834 roLeuLleGlySerAsnAsnValLeuLeuGlnGluMetAspThrTyrGlu 850
2501 CCTTATTCGGAGCAACAATGCTCTCTGGAGAGATGAGACACTATGAG 2550
851 TrpAlaLeuLysSerTrpAlaProCysSerLysAlaCysGlyGlyLyl 867
2551 TGGGGCTTCAMAGAGCTGGGCCCTGCAGCAAGGCTGTGGAGAGGAT 2600
867 eGlnPheThrLysTyrGlyCysArgArgArgArgAspHisHisMetValG 884
2601 CCAAGTTCACCAATAGCGCTCCGGCGCAGACGAGACACCAATGTTGTC 2650
884 lInArgHisLeuCysAspHisLysLysArgProLysProIleArgArgArg 900
2651 AGCGACACCTGTGTGACCAACAAGAAAGAGCCCAAGCCATCCGCGGCGC 2700
901 CysAsnGlnHisProCysSerGlnProValTrpValThrGlnLutrrpGl 917
2701 TCACAACAGCACCCGCTGCTCAGCCTGTGGGTGACGAGAGAGTGGGG 2750
917 yAlaCysSerArgSerCysGlyLysLeuGlyValGlnThrArgGlyLleG 934
2751 TGCTTCGACGGGAGACTGTGGAAAGCTGGGGGTGCACACAGGGGGATAC 2800
934 lnCysLeuLeuProLeuSerAsnGlyThrHisLysValMetProAlaLys 950
2801 AGTGCCTGCTGCCCTCTCCAAATGAAACCCCAAGATGATCGCGGCATA 2850
951 AlaCysAlaGlyAspArgProGluAlaArgArgProCysLeuArgValTr 967
2851 GCGGTGCGGGGACCGGCTGAGAGCCGACGGCCCTGCTCCGAGTGGCC 2900
967 oCysProAlaGlnTrpArgLeuGlyAlaTrpSerGlnCysSerAlaTrrhc 984
2901 CTGCCAGACCCAGTGAAGGCTGGGAGGCTGTGCCAGTGTCTGCCACCT 2950
984 ysglyGlyGlyLleGlnGlnArgGlnValValCysArgThrAsnAlaAsn 1000
2951 GTGAAGAGGGGATCCACAGCGGAGAGTGTGTGCAGGACCAACGCCAAC 3000
1001 SerLeuGlnHisCysGluGlyAspArgProAspThrValGlnValCysSe 1017
3001 AGCCTCGGCGCATTCGGAGGGGATAGGCCAGACACTGTCCAGGTCTGCAG 3050
1017 rLeuProAlaCysGlyLysAsnHisGlnAsnSerThrValArgAlaAspV 1034
3051 CCTGCTGCTGTGGAGAAATCACCAAGATCCACAGGTGAGAGGCGGAGT 3100
1034 alrrpGluLeuGlyLyrProGluGlyGlnTrrpAlProLinsErGlyPro 1050
3101 TCTGGAGACTTTGGAGCCAGAGAGGGCAAGTGGTGCACAAATCTTAACCC 3150
1051 LeuHisProIleAsnLysLleSerSerMetCysAlaAlaGlnProCysTrh 1067
3151 CTACATCCCATTAACAAGATATCATCA.....ACGAGACCTGTGCAC 3191
1067 rGlyAspArgSerValPheCysGlnMetGluValLeuAspArgTyrCysS 1084
|||||

```

3192 GGGAGACAGTCTGTCTTTCGACATGAAAGTCTGATGCTACTGCT 3241  
 1084 erileprogllythrhleuargleucysvalsercysileylsalsala 1100  
 3242 CCAATCCCGGCTACACACGGGCTGTGTGTCTGATCAGAAAGGCC 3291  
 1101 SerGlyProAsnProGlyProAspProGlyProThrSerLeuProProh 1117  
 3292 TCGGGCCCAACCTGGCCACGACCTGGCCCAACCTGCTGCCCCCTT 3341  
 1117 eserThrProGlySerProleuProGlyProGlnAspProAlaAspAla 1134  
 3342 CTCACGCTCCGGAAGCCCTTACAGAGACCCAGAGACCTGCGATGCTG 3391  
 1134 laGluProProGlyLysProThrGlySerGluAspHisGlnHisGlyArg 1150  
 3392 CAGAGCTCCCTGGAAAGCAAGGATCAGAGACCATCAGCATGGCCGA 3441  
 1151 AlaThrGlnLeuProGlyAlaLeuAspThrSerSerProGlyThrGlnH 1167  
 3442 GCCACACAGCTCCAGAGACTGTGATTAACAAGCTCCCAAGGACCCAGCA 3491  
 1167 sProPheAlaProGluThrProIleProGlyAlaSerTrpSerIleSer 1184  
 3492 TCCCTTTCCTCCCTGAGACACCAATCCTGGAGCATCTGGAGCATCTCC 3541  
 1184 rothrThrProGlyLysLeuProTrpGlyTrpThrGlnThrProThrPro 1200  
 3542 CTACACACCCCGGGGGGCTGCTGGGGCTGAGCTCAGACACCTACGCCA 3591  
 1201 ValProGluAspLysGlyGlnProGlyLysLeuAspLeuArgHisProGlyTh 1217  
 3592 GTCCCTGGAGCAAAAGGCAACCTGGAGAGACCTGAGACATCCCGGCAC 3641  
 1217 rSerLeuProAlaAspLeuPro 1224  
 3642 CAGCTCTCCCTGTGCTGCCCG 3663

seq\_name: gb-pat:AX319856

seq\_documentation\_block: 3561 bp DNA linear PAT 14-DEC-2001  
 LOCUS AX319856  
 DEFINITION Sequence 20 from Patent WO0183782.  
 ACCESSION AX319856  
 VERSION AX319856.1 GI:17901446

KEYWORDS

SOURCE

human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (sites)  
 Plowman, G.D., Whyte, D., Sudarsanam, S., Manning, G., Caenepeel, S. and

OTHERS

Patent: WO 0183782-A 20 08-NOV-2001

FEATURES

Location/Qualifiers  
 1..3561  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 694 a 1131 c 1113 g 623 t  
 ORIGIN

alignment\_scores:  
 Quality: 6508.00 Length: 1224  
 Ratio: 5.511 Gaps: 1  
 Percent Similarity: 96.487 Percent Identity: 96.324

alignment\_block:

US-09-938-330-20 x AX319856 ..

Align seg 1/1 to: AX319856 from: 1 to: 3561

1 MetAlaProLeuArgAlaLeuLeuSerTyrLeuLeuProLeuHisCysAl 17  
 1 ATGGCTCCACCTCGCGCTGTCTCTACTGCTGCTTGGCTTGGCTGGC 50  
 17 aleuCystrAlaAlaGlySerArgThrProGluLeuHisLeuSerGlyL 34  
 51 GCTTCGCCCGCGCGGGCAGCGGACCCCAAGAGCTGCACCTCTCTGGA 100  
 34 yaleuSerAspTyrGlyValThrValProCysSerThrAspPheArgGly 50  
 101 ACCTCAGTCACTATGTGTGACAGTGCCTCAGCAGCACTCTTGCGGGA 150  
 51 ArgPheLeuSerHisValSerGlyProAlaAlaAlaSerAlaGlySe 67  
 151 GCCTTCTCTCCACGCTGTGTCTGTGCCCAGCAGCAGCCTCTGCGAGG 200  
 67 rMetValAlaAspThrProProThrLeuProArgHisSerSerHisLeu 84  
 201 CATGGTAGTGACAGCCACCACACTACACAGACACTCCACTCCCTCC 250  
 84 rGValAlaArgSerProLeuHisProGlyGlyThrLeuTrpProGlyArg 100  
 251 GGGTGGCTGCGAGCCCTGTGACCCAGAGGAGCCCTGTGGCTGGCAGG 300  
 101 ValGlyArgHisSerLeuTyrPheAsnValThrValPheGlyLysGlu 117  
 301 GTGGGGCGGCACCTCCCTCACTCAATGTCACTGTTCGGGAAGCAACT 350  
 117 whisLeuArgLeuArgProAsnArgArgLeuValValProGlySerSer 134  
 351 GCACCTTGCGCTGGCGCCCAATCGGAGGTGGTAGTGCCAGGATCCAG 400  
 134 alGluTrpGlnGluAspPheArgGluLeuPheArgGlnProLeuArgGln 150  
 401 TGGAGTGGCAGAGAGATTTTCGGGAGCTGTCCGGAGGCCCTTAAGGCG 450  
 451 GAGTGTGTGACACTGAGGTGTCACTGGAATGCCGGGGCGAGCTGTTC 500  
 151 GluCyValTyrThrGlyLysValThrGlyMetProGlyAlaAlaValAl 167  
 167 alIleSerAsnCysAspGlyLeuAlaGlyLeuLeuArgThrAspSerTh 184  
 501 CATCAGCAACTGTGACGAGATGGCGGCTCATCCCAAGACAGCAGCAGC 550  
 184 sPhePheIleGluProLeuGluArgGlyGlnGlnLeuLysGluAlaSer 200  
 551 ACTTCTTCATTTAGCCTCTGAGCGGGGCCAGCAGAGAAGGCGCAGC 600  
 201 GlyArgThrHisValValTyrArgArgGluAlaValGlnGlnGluTrpAl 217  
 601 GGGAGGACACATGTGTGTACCCCGGGAGGCCGTCACACAGAGAGTGGC 650  
 217 agluProAspGlyAspLeuHisAsnGlnAlaPheGlyLeuGlyAspLeu 234  
 651 AGAACCTGACGGGAGCCTGCACAATGAAGCTTTGGCGTGGAGAGACTTC 700  
 234 roAsnLeuLeuGlyLeuValGlyAspGlnLeuGlyAspThrGluArgLys 250  
 701 CCAACCTGTGGGCTGTGGGAGACAGCTGGGCGACAGAGAGCGGAAG 750  
 251 ArgArgHisAlaLysProGlySerTyrSerIleGluValLeuLeuVal 267  
 751 CGGGGCGATGCCAAGCCAGGACCTCAGCATCAGATGAGTGTGTGTGT 800  
 267 laspAspSerValValAlaPheHisGlyLysGluHisValGlnAsnTyr 284  
 801 GGAGAGACTCGGTGTGTCTTCATGCAAGAGAGCATGTGTCAAACTATG 850  
 284 alLeuThrLeuMetAsnIleValAspGluIleTyrHisAspGluSerLeu 300  
 851 TCTCACCCCTCATGAATATCTGATGATGATTTACACGATGATGCTCC 900

```

301 GlyValHisIleasnIlealeuValArgLeuIleMetValGlyTyr 317
    |||||||
901 GGGGTTCAATATAATATTCCTCCCTCCGCTTGATCATGTGGCTACG 950
    |||||||
317 gGlnSerLeuSerLeuIleGlyArgGlyAsnProSerArgSerLeuG 334
    |||||||
951 ACAGTCCCGAGCCCTGATGAGAGCGGGAGACCCCTCACGAGCCTGGAG 1000
    |||||||
334 lValCysArgTrpAlaHisSerGlnIleArgGlnAspProSerHisAla 350
    |||||||
1001 AGGTGTGTCCTGGGACACCTCCAGCGCGCAGGACCCGACGAGCT 1050
    |||||||
351 GluHisHisAspHisValValPheLeuThrArgGlnAspPheGlyPro 367
    |||||||
1051 GAGCACCATGACACCGTTGTGTCTCTACCCGGCAGGACTTTGGCCCTC 1100
    |||||||
367 rGlyTyrAlaProValIleThrGlyMetCysHisProLeuArgSerCysAla 384
    |||||||
1101 AGGGTATGACCCCGTCACTGGCATGTGTACACCCCTGAGAGCTGTGCC 1150
    |||||||
384 euAsnHisGluAspGlyPheSerSerAlaPheValIleAlaHisGluThr 400
    |||||||
1151 TCACCATGAGAGATGGCTTCTCCACGCTTGATATAGCTCATGAGAGC 1200
    |||||||
401 GlyHisValLeuGlyMetGluHisAspGlyGlnGlyAsnGlyCysAla 417
    |||||||
1201 GGCCACGCTGCTGGCATGAGCATGAGCGGTCAAGGAAATGGCTGTGAGA 1250
    |||||||
417 pGluThrSerLeuGlySerValIleAlaProLeuValGlnAlaAlaPhe 434
    |||||||
1251 TGAGACGAGCCTGGGACAGCTCATAGCGCCCTGTGTGACGCTGCCCTCC 1300
    |||||||
434 lAspArgPheHisTrpSerArgCysSerLysLeuGluLeuSerArgTyrLeu 450
    |||||||
1301 ACCGTTTCATTGGTCCCGCTGCAGCAAGTGGAGCTCACGCCCTACCTC 1350
    |||||||
451 ProSerTyrAspCysLeuLeuAspAspProPheAspProAlaTrpProG 1400
    |||||||
1351 CCTCTCTAGGACTGCTCTGTGATGACCCCTTGTATCTCTCCGCGCCCA 1450
    |||||||
467 nProProGluLeuProGlyIleAsnTyrSerMetAspGlyGlnCysArgP 484
    |||||||
1401 GCCCCAGAGCTGCCGTGGATCAACTCAATGATGAGACATGGCGCT 1450
    |||||||
484 heAspPheGlySerGlyTyrGlnThrCysLeuAlaPheArgThrPheGlu 500
    |||||||
1451 TTGACTTGGCAGTGGCTACACGACCTGTGGCATTCAGAGACCTTTGAG 1500
    |||||||
501 ProCysLysGlnLeuTrpCysSerHisProAspAsnProTyrPheCys 517
    |||||||
1501 CCTGCAAGCAGCTGTGTGCTGACGACCATCTGACACACCCGACTTGTGCA 1550
    |||||||
517 sThrLysLysGlyProProLeuAspGlyThrGlyCysAlaProGlyLys 534
    |||||||
1551 GACCAAGAGAGGGGCCCTGGATGGAGTGAAGTGTGCACCCGGCAAT 1600
    |||||||
534 rPysPheLysGlyHisCysIleTrpLysSerProGluGlnThrTyrGly 550
    |||||||
1601 GGTCCTCAAGGTCATGTCATGTGAAGTCGCGGAGCAGACATATGAC 1650
    |||||||
551 GluAspGlyLysTrpSerSerTrpThrLysPheGlySerCysSerArg 567
    |||||||
1651 CAGGATGAGAGCTGGAGCTCTGACCAAGATTGGGTCTGTTCGGGGTCC 1700
    |||||||
567 rCysGlyGlyValAlaArgSerArgSerArgSerCysAsnAsnProSer 584
    |||||||
1701 ATGTGGGGGGGGGTGGATCCCGCAGCGGAGCTCAACAACCCCTGCC 1750
    |||||||
584 roAlaTyrGlyGlyArgProCysLeuGlyProMetPheGlyTyrGlnVal 600
    |||||||
1751 CAGCCTATGAGGCCGCCGTGTGTAGGGCCCATGTTCAGTACCAAGGTC 1800
    |||||||
601 CysAsnSerGlnGluCysProGlyThrTyrGluAspPheArgAlaGln 617
    |||||||
    |||||||
1801 TCACAGAGCAGAGAGTGGCTGGGACCTACGAGAGCTCCGGCCAGCA 1850
    |||||||
617 nCysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHisSer 634
    |||||||
1851 GTGTGCCAAGCCCAACTGCTATGTGTACCAAGAAATGCCAAGCAGACT 1900
    |||||||
634 rPValProTyrGluProAspAspAlaGlnLysCysGluLeuIleCys 650
    |||||||
1901 GCGTCCCTACAGAGCCTGACGATGACGCCCAAGAGTGTGAGCTATCTGC 1950
    |||||||
651 GluSerAlaAspThrGlyAspValValPheMetAsnGlnValValHis 667
    |||||||
1951 CAGTGGGGGACACGGGGAGCTGTGTTCATGAACCAAGGTGGTTCCAGA 2000
    |||||||
667 pGlyThrArgCysSerTyrArgAspProTyrSerValCysAlaArgGly 684
    |||||||
2001 TGGACACGCTGACACTACCGGACCCATACAGCGCTGTGTGGCGGTGCG 2050
    |||||||
684 lCysValProValGlyCysAspLysGluValGlySerMetLysAlaAsp 700
    |||||||
2051 AGTGTGCTGCTGGCTGCTGACAAAGAGGTGGGGTCCATGAAGCGGAT 2100
    |||||||
701 AspLysCysGlyValCysGlyGlyAspAsnSerHisCysArgThrVal 717
    |||||||
2101 GACAACTGTGAGTGTGGGGGTGACACCTCCACCTGACGAGACTGTGAA 2150
    |||||||
717 sGlyThrLeuGlyLysAlaSerLysGlnAlaGlyAlaLeuLysLeuVal 734
    |||||||
2151 GGGGAGCTGGGCAAGGCTCCAAAGCAGGACGAGCTCTCAACTGTGTG 2200
    |||||||
734 lIleProAlaGlyAlaArgHisIleGlnIleGlyAlaLeuGlyLysSer 750
    |||||||
2201 AGATCCACAGCAGGTGCCAGGACATCCAGATTGAGGACACGAGAACTGC 2250
    |||||||
751 ProHisArgSerValValLysAsnGlnValThrGlySerPheIleLeu 767
    |||||||
2251 CCCACCGCATTTGTGTGAAGAACCAGTACCGGACGCTTCACTCTCAA 2300
    |||||||
767 nProLysGlyLysGluAlaThrSerArgThrPheThrAlaMetGlyLeu 784
    |||||||
2301 CCCCAAGGGCAAGGAAAGCCCAAGCCGACCTTACCCGCTGAGGCTGG 2350
    |||||||
784 lIleTrpGluAspAlaValGluAspAlaLysGluSerLeuLysTrpSer 800
    |||||||
2351 AGTGGAGAGATGCCGTGAGAGATGCCAAGAAAGCTCTCAAGACACGGGG 2400
    |||||||
801 ProLeuProGluAlaIleAlaIleLeuAlaLeuProProThrGluGly 817
    |||||||
2401 CCTGTGCTGAAGCATTTGCCATCTGTGCTTCCCTCCCAACTGAGGGTGG 2450
    |||||||
817 yProArgSerSerLeuAlaTyrTrpTyrValIleHisGluAspLeuArg 834
    |||||||
2451 CCCCCGAGAGCCTGGCTGACCAAGTACGTATCCATCGAGGAGCTGTGTC 2500
    |||||||
834 roLeuIleGlySerAsnAsnValLeuLeuGluGluMetAspThrTyrGlu 850
    |||||||
2501 CCTTATCGGAGACCAACAAATGTCTCTGTGAGAGAGATGACACCTATGAG 2550
    |||||||
851 TrpAlaLeuLysSerTrpAlaProCysSerLysAlaCysGlyGlyLys 867
    |||||||
2551 TGGCGCTCAAGAGCTGGGGCCCTGACCAAGGCTGTGAGGAGGAGAT 2600
    |||||||
867 eGlnPheThrLysTyrGlyCysArgArgArgArgAspHisHisMetVal 884
    |||||||
2601 CCACTTACCAAAATACGGCTGGCGGGGACAGAGACCAACCAATGAGTGC 2650
    |||||||
884 lAspHisLeuCysAspHisLysLysArgProLysProIleArgArgArg 900
    |||||||
2651 AGCGACACCTGTGTGACCAAGAAAGGCCCAAGCCATCCGCGGGGCG 2700
    |||||||
901 CysAsnGlnHisProCysSerGlnProValIleThrGluGluTrpGly 917
    |||||||

```

2701 TGCACACAGCACCCTGCTCTCAGCCCTGTGGGTGACGAGAGTGGG 2750  
 917 YALACYSERARGSERCYSGLYLSEUGLYVALGINTRARGLYL 934  
 2751 TGCCTGACGGGAGCTGTGGAAAGCTGGGGGTGACACAGGGGGATAC 2800  
 934 INCYSLEULEUPROLEUSERASNGLYTHRHSLYSVALMETPROALAYS 950  
 2801 AGAGCCGATGCCCCCTCTCCATGAAACCCACAAAGGTCAATGCCGCCAAA 2850  
 951 ALACYSALAGLYSPARGPROGLUALAARGARGPROCYSEUNARGVALP 967  
 2851 GCCCTGCTCCGGGACCGGCTGAGGCGCCGAGCCCTGTCTCCGAGTGCC 2900  
 967 OCYSPROALAGINTRPARGLAVALATPSPERGINCYSERVALTHC 984  
 2901 CTGCCACGCCACGAGAGGCTGGAGGCTGTCCAGTGTCTGCCACCT 2950  
 984 YSGLYGLUGLYLEGLINARGGLNVALCYSPARGTHRASNALAASN 1000  
 2951 GTGAGAGGCGATCCACAGCGGAGGTGTGACAGACCAACGCCAAC 3000  
 1001 SERLEUGLYHSCYSGLYLSPARGPROASPPTHVALGINVALCYSE 1017  
 3001 AGCCTCGGCGATGCGAGGGGATAGGCCACACTCTCCAGGTCTGAC 3050  
 1017 RLEUPROALACYSGLYLSNHSINHSNHSERTHVALARGALASP 1034  
 3051 CCTGCTGCTGTGGA..... 3066  
 1034 ALTPGLULEUGLYTHRPROGLUGLYNTRPVALPROGLINSERGLY 1050  
 3066 ..... 3066  
 1051 LEUHSIROLEASNLYSLESERSEMETCYSAALAGLUPROCYSTH 1067  
 3067 .....GCGGAGCCCTGCAC 3080  
 1067 RGLYASPARSERVALPHECYSGINMETGLUVALLEUASPARGYTCSS 1084  
 3081 GGGAGACAGGTGTGCTCTGCCAGATGAGATCTCGATCGCTACGTGCT 3130  
 1084 ERLLEPROGLYTHRHSARGLEUCYCYSALESERCYSLLELYSALA 1100  
 3131 CCATTCGCCGCTACACCGGCTGTGTGTCTGTCTGATCAAGAGGCC 3180  
 1101 SERGLYPROASNPROGLYPROASPARGLYPROTHRSERLEUPROPH 1117  
 3181 TCGGGCCCCAACCCCTGGCCAGACCCCTGCCAACCTCACCTGCCCTT 3230  
 1117 ESERTHPROGLYSERPROLEUPROGLYPROGLINASPVALAASPVAL 1134  
 3231 CTCCACTCTGAGAGCCCTTACACAGGACCCACAGACCCCTGACATGCTG 3280  
 1134 LAGLUPROPROGLYLYSPROTHRGLYSERGLUASPHISGLINISGLYARG 1150  
 3281 CAGAGCTCTTGAAGAACCAAGGATGAGAGACATCAGCATGCGCA 3330  
 1151 ALATHRGINLEUPROGLYVALALEASEPHRSERSEPROGLYTHGNHI 1167  
 3331 GCCACACAGCTCCAGAGACTGTGATACAGCTCCCAAGGACCCACAGA 3380  
 1167 SPROHEALAPROGLUTHRPROILEPROGLYVALASERTPSERISE 1184  
 3381 TCCCTTGGCCCTGAGACACCAATCCCTGAGCATCTGAGACATCTCC 3430  
 1184 ROTHTHRPROGLYLYLEUPROTPRGLYTPRPHGLINTHRPROTHR 1200  
 3431 CTACACACCCCGGGGGCTGTGGGCTGAGCTCAGACACCTACGCCA 3480  
 1201 VALPROGLUASPLYSGLYINPROGLYLUASPLUARGHISPRGLYTH 1217  
 3481 GTCCCTGAGAGCAAGGCAACCTGAGAGAACCTGAGACATCCCGGCAC 3530

1217 rserleuproalasprou 1224  
 3531 CAGCCTCCCTGCTGCTCCCG 3552  
 seq\_name: gb\_pat:AX342634

seq\_documentation\_block: 3567 bp DNA linear PAT 12-JAN-2002  
 LOCUS AX342634  
 DEFINITION Sequence 31 from Patent WO0198468.  
 ACCESSION AX342634  
 VERSION AX342634.1 GI:18152031  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 (sites)  
 Yue, H., Elliott, V. S., Gandhi, A. R., Lal, P., Au-Young, J.,  
 Tribouley, C. M., Delegeane, A. M., Baughn, M. R., Nguyen, D. B., Lee, E. A.,  
 Hafalia, A., Khan, F. A., Walla, N. K., Yao, M. G., Lu, D. A., Patterson, C.,  
 Tang, J. T., Walsh, R. T., Azimzai, Y., Rankumar, J., Xu, Y. and Reddy, R.  
 Patent: WO 0198468-A 3 (27-DEC-2001)

JOURNAL  
 Incyte Genomics, Inc. (US)  
 location/Qualifiers  
 1..3567  
 /organism="Homo sapiens"  
 /dp\_xref="taxon:9606"  
 /note="Incyte ID No: 7477386CB1"

BASE COUNT 692 a 1133 c 1117 g 625 t  
 ORIGIN

alignment\_scores:  
 Quality: 6468.50 Length: 1227  
 Ratio: 5.482 Gaps: 4  
 Percent Similarity: 96.170 Percent Identity: 96.088

alignment\_block:  
 US-09-938-330-20 x AX342634 ..

Align seg 1/1 to: AX342634 from: 1 to: 3567

1 MetAlaProLeuArgAlaLeuLeuSerTYRLeuLeuProLeuHISCYAL 17  
 1 ATGGCTCCACTCCGCGCTGTCTCTACCTGCTGCTCTTCTCAGCTGTC 50  
 17 ALAUCYSPTHALALAGLYSERARGTHRPROGLULEUHSLEUSERGLY 34  
 51 GCTCTGGCGCGCGCGGAGCGGAGCCAGAGCTGACACTCTGGGA 100  
 34 YLSLEUSERTYRGLYVALTHVALPROCYSETRHASPHEARGLY 50  
 101 AGCTCAGTGACTATGGTGTGACAGTCCCTGACAGCACTTTCCGGGA 150  
 51 ARGPHLEUSERHISVALVALSERGLYPROALAALASERVALYSE 67  
 151 CGCTCTCTCCACAGTGTGTGTGGCCAGACAGCACTCTGACAGGAG 200  
 67 IMETVALVALASPTHRPROBOTHRLEUPROARGHISSESERHISLEUA 84  
 201 CATGTGTGTGACACGCCACCACTACACAGACACTCCAGTCACTCC 250  
 84 RYVALALADGSETRPROLEUHSISPROGLYLYTHRLEUPRPROGLYARG 100  
 251 GCGTGTCTCCAGCCCTCTGACCCAGAGAGGAGCCCTGTGGCTGACAG 300  
 101 VALGLYARGHISSETRLEUTYRPHASNVALTHVALPHEGLYLSGLULE 117  
 301 GTGGGGCGCACCTCCCTTCAATGATGACACTTTTGGGAAGGAACT 350  
 117 UHSLEUARGLEUARGPROASNARGARGLEUVALVALPROGLYSESERV 134

```

351 GCACTTGCGCTGGGCGCCCAATCGAGGTTGGTATGCCAGGATCTCTCAG 400
134 ALGIUTPGLINGLUAAspPheArgGLuleuPheArgGLInProleuArgGLIn 150
401 TGGAGTGTGCAGAGAGATTTTCGGAGCTGTCTCGGCGACGCTTACGGCAG 450
151 GLUCysValTyrThrGLysValAlaThrGLyMetProGLyAlaAlaValAla 167
451 GAGGTGTGTACACTGTGAGGTGTACACTGCAATGCTGGGGCAGCTGTGC 500
167 ALIeserAsnCysAspGLyLeuAlaGLyLeuIleArgThrAspSerThrA 184
501 CATCAGCAACTGTGACGGAATGGCGGGCTCATCCGACAGACAGACACCG 550
184 sPhePheIleGIuProleuGIuArgGLyGLInGLInGLyLysGLyLys 200
551 ACTTCTTCATTGAGCTCTGAGCGGGCGCAGCAGAGAGAGAGGCGCAGC 600
201 GLYArgThrHisValValTyrArgArgGLuAlaValGLInGLInGLyTrpAl 217
601 GGGAGGACACATGTGTGTACCGCGGGAGCGCTCCAGCAGAGATGGGC 650
217 AGIuProAspGLyAspLeuHisAsnGLuAlaPheGLyLeuGLyAspLeuP 234
651 AGAACTGACGGGAGCTGCACAAATGAACCTTTGGCTGGGAGACCTTC 700
234 roAsnLeuLeuGLyLeuValGLyAspGLInLeuGLyAspThrGLInArgLys 250
701 CCMACTGCTGGGCGCTGGTGGGGGACCACTGGCGACACACAGGGGAG 750
251 ArgArgHisAlaLysProGLySerTyrSerIleGLyValLeuLeuValVa 267
751 CGCGGCAATGCCAGGCCAGCGACGCTACAGCATCGAGTGTCTGTGTGT 800
267 LAspAspSerValValArgPheHisGLyLysGLInHisValGlnAsnTyrV 284
801 GGACGACTGGTGGTTCCTTCATGTGCAGAGAGCATGTGCAGAACTATG 850
284 ALIeuThrLeuMetAsnIle...ValAspGLyIleTyrHisAspGLySer 299
851 TCCTCACTCATGAATATCGTGTAGATGAGATTACACAGATGAGATCC 900
300 LeuGLyValHisIleAsnIleAlaLeuValArgLeuIleMetValGLyTyr 316
901 CTGGGGGTTCAATAAATATGCTCGCTCGCTGATCATGATGGTGGCTA 950
316 rArg...GlnSerLeuSerLeuIleGLyArgGLyAsnProSerArgSerL 332
951 CGGACAGCAGTCCCTGAGCTGTATCGAGCGGGAGACCCCTCACGCGACC 1000
332 euGLInValLysArgTyrPheHisSerGLInArgGLInAspProSer 348
1001 TGGACAGAGTGTGTGGCGGACACTGCCAGCGCCGAGAGACCCAGC 1050
349 HisAlaGLInHisAspHisValValPheLeuThrArgGLInAspPheGL 365
1051 CAGGCTGAGCAACATGACGACGAGTGTGTCTCTACCCGCGAGGACTTGG 1100
365 yProSer...GLyTyrAlaProValThrGLyMetCysHisProleuArgS 381
1101 GCCCTCAGGTGGTATGACCCGCTCATCTGGATGTGTACCCCTTGAGGA 1150
381 eCYsAlaLeuAsnHisGLInAspGLyPheSerSerAlaPheValIleAla 397
1151 GCTGTGCCCTCAACCATGAGGATGGCTTCTCCACAGCTTCGTGATAGCT 1200
398 HisGLuThrGLyHisValLeuGLyMetGLInHisAspGLyGLInLysAsn 414
1201 CATGAGACCGGCGCACTGCTCTGGCATGAGCATATGACGGTACGGGATGG 1250
414 yCYsAlaAspGLyThrSerLeuGLySerValMetAlaProLeuValGLInA 431
1251 CTGTGACAGATGAGACAGCCTGGGCGACGTCATGGCGCCCTGTGTACAG 1300
431 IAlaPheHisArgPheHisTrpSerArgCysSerLysLeuGLuleuSer 447
1301 CTGGCTTCCACCGCTTCATTTGGTCCCGCTGCAGCAGATGGAGCTCAGC 1350
448 ArgTyrLeuProSerTyrAspCysLeuLeuAspAspProPheAspProAl 464
1351 CGCTACTCCCGCTCTACAGACTGCTCTCGATGACCCCTTTGATCTCTGC 1400
464 aTrpProGLInProGLyLeuLeuPheGLyIleAsnTyrSerMetAspGLu 481
1401 CTGGCCCGACCGCCCGACAGCTGCTGGATCACTCAATCAATGATGAGC 1450
481 IncysArgPheAspPheGLySerGLyTyrGLInThrCysLeuAlaPheArg 497
1451 AGTGGCGCTTGTGACTTGTGACAGTGGCTACACAGACCTGCTGGCATTCAG 1500
498 ThrPheGLuProCysLysGLInLeuTrpCysSerHisProAspAsnProTy 514
1501 ACCTTGTAGCCCTGCAGAGCAGCTGTGTGACAGCCATCCTGACAAACCGTA 1550
514 rPheCysLysThrLysLysGLyProProleuAspGLyThrGLyCysAlaP 531
1551 CTCTCTCAAGACCAAGAAAGGGCCCGCTGATGGAGCTAGATGTGTAC 1600
531 roGLyLysTrpCysPheLysGLyHisCysIleTyrPheLysSerProGLuGL 547
1601 CCGGCAAGTGTGTCTCAAGGTCACCTCATCTGTGAAGTCCCGAGACAG 1650
548 ThrTyrGLyLysAspGLyGLyTrpSerSerTrpThrLysPheGLySerCy 564
1651 ACATATGGCCAGAGATGAGAGCTGGAGCTCTGGACCAAGTTGGGATATG 1700
564 sSerArgSerCysGLyGLyValArgSerArgSerArgSerCysAsnA 581
1701 TTGGGGTATGTGGGGGGGGGGTGCATCCCGACGCGGAGCTGCACA 1750
581 snProSerProAlaTyrGLyArgProCysLeuGLyPrometPheGLu 597
1751 ACCCTCGCCAGCTATGAGGCGCCCTGTCTTAAAGGCCCATCTTTCAG 1800
598 TyrGLInValCysAsnSerGLyLysCysProGLyThrTyrGLInAspPheAr 614
1801 TACAGAGTCTGCACAGCAGAGAGATGCCCTGGAGCTTACGAGACTTCGC 1850
614 gAlaGLInCysAlaLysArgAsnSerTyrTyrValHisGLInAsnAlaL 631
1851 GGCCGAGCAGTGTGCCAAGCGCAACTCTCTACTATGTGCACAGAAATGCCA 1900
631 ySHISerTrpValProTyrGLuProAspAspAlaGLInLysCysGLu 647
1901 AGCAGAGTGGGTGCCCTACAGGCTTACAGATGACAGCCCGAAGTGTGAG 1950
648 LeuIleCysGLInSerAlaAspThrGLyAspValValPheMetAsnGLInV 664
1951 CTGATCTGCCAGTGGCGGAGACGCGGGGAGCTGTGTCTTGAACCAAGGT 2000
664 lValHisAspGLyThrArgCysSerTyrArgAspProTyrSerValCysA 681
2001 GATTACAGATGGGACAGCGCTGCACGCTACCGGAGCCATACAGCGCTGTG 2050
681 lAArgGLyLysCysValProValGLyCysAspLysGLuValGLySerMet 697
2051 CGCGTGGCGAGTGTGTCTCTGTGGCTGTACAAAGAGGTGGGGTCCAGT 2100
698 LysAlaAspAspLysCysGLyValLysGLyLysAspAsnSerHisCysAr 714
2101 AAGCGGATGACAGTGTGAGTCTGGGGGGGTGACAACTCCACCTCGCAG 2150
714 gThrValLysGLyThrLeuGLyLysAlaSerLysGLInAlaGLyAlaLeu 731
2151 GACTGTGAAGGGGAGCGCTGGGCAAGGCTTCCAAAGCAGGAGGTGTCTCA 2200

```



731 ysluValGlnIleProAlaGlyAlaArgHisIleGlnIleGlnAlaLeu 747  
 2201 AGCTGTGTCAGATCCACAGCAGTGCACATCCAGATGAGGCACTG 2250  
 748 GluYssSerProHisArgSerValValLysnGlnValThrGlySerP 764  
 2251 GAGAGTCCCTCCACGATTGTGTGAAGACCAAGGTCACCGGAGCTT 2300  
 764 eileuAsnProLysGlyLysGlnAlaThrSerArgThrPheThrAla 781  
 2301 CATCTTCACCCCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 2350  
 781 etGlyLeuGlnTrpGlnAspAlaValGluAspAlaLysGlnSerLeu 797  
 2351 TGGGCTGGAGTGGAGGATGGGAGGATGGGAGGATGGGAGGATGG 2400  
 798 ThrSerGlyProLeuProGlnAlaIleAlaIleAlaLeuProPhe 814  
 2401 ACCAGCGGCGCTCTGCTGAGGACCTTCATCTGCTGCTGCTGCT 2450  
 814 rGluGlyLysProArgSerSerLeuAlaTyrLysTyrValIleHisG 831  
 2451 TGAGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2500  
 831 sPLeuLeuProLeuIleGlySerAsnAsnValLeuLeuGlnGluMet 847  
 2501 ACCCTGCTGCTCTTATCGGAGCAACATGTCTCTGAGGAGATGG 2550  
 848 ThrTyrGlnTrpAlaLeuLysSerTrpAlaProCysSerLysAla 864  
 2551 ACCCTATGAGTGGGCGCTCAAGAGCTGGGCGGCGGCGGCGGCG 2600  
 864 yGlyGlyIleGlnPheThrLysTyrGlyCysArgArgArgArgAs 881  
 2601 AGGAGTATCCAGTTCACCAATACGGCTGGCGGCGGAGACAGACC 2650  
 881 iMetValGlnArgHisLeuCysAspHisLysLysArgProLysPro 897  
 2651 ACATGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2700  
 898 ArgArgArgCysAsnGlnHisProCysSerGlnProValThrVal 914  
 2701 CGCGGAGCTGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2750  
 914 uGlnTrpGlyAlaCysSerArgSerCysGlyLysLeuGlnValIle 931  
 2751 GAGGTGGGCTGCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAG 2800  
 931 rGlyIleGlnCysLeuLeuProLeuSerAsnGlyThrHisLysVal 947  
 2801 GGGGATACAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2850  
 948 ProAlaLysAlaCysAlaGlyAspArgProGlnAlaArgArgPro 964  
 2851 CCGGCAAGGCTGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 2900  
 964 uArgValProCysProAlaGlnTrpArgLeuGlyAlaTrpSerGln 981  
 2901 CCGAGTGCCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2950  
 981 eAlaThrCysGlyGlyIleGlnGlnArgGlnValValCysArgThr 997  
 2951 CTGCCACCTGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3000  
 998 AsnAlaAsnSerLeuGlyHisCysGlyLysAspArgProAspThr 1014  
 3001 AAGGCCAAGCCTCGGCAATTGCGAGGAGGAGGAGGAGGAGGAG 3050  
 1014 nValCysSerLeuProAlaCysGlyLysnHisGlnAsnSerThrVal 1031  
 3051 GGTCTGAGGCTGCGGCGGCTGTGA..... 3075  
 1031 rGlnAspValTrpGluLeuGlyThrProGluGlyLysTrpValPro 1047

3075 ..... 3075  
 1048 SerGlyProLeuHisProIleAsnLysIleSerSerMetCysAlaAla 1064  
 3076 ..... GCGGA 3080  
 1064 uProCysThrGlyAspArgSerValPheCysGlnMetGluValLeuAsp 1081  
 3081 GCCCTGACGAGGAGACAGCTGTCTTCTTCCAGATGAGTAAGTCTG 3130  
 1081 rGlyCysSerIleProGlyTyrHisArgLeuCysCysValSerCysIle 1097  
 3131 GCTAGTCTCCATTCCTCGGCTACACCGGCTGCTGTGTCTGCTGCA 3180  
 1098 LysLysAlaSerGlyProAsnProGlyProAspProGlyProThr 1114  
 3181 AGAAGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3230  
 1114 uProProPheSerThrProGlySerProLeuProGlyProGlnAsp 1131  
 3231 GCCCTCTTCTCCATCTCTGAGAGCCCTTACAGAGGAGGAGGAG 3280  
 1131 LaAspAlaAlaGluProProGlyLysProThrGlySerGlnAspHis 1147  
 3281 CAGATGCTGCGAGCTCTCTGAAAGCAAGGATCAGAGGACCAT 3330  
 1148 HisGlyValArgAlaThrGlnLeuProGlyAlaLeuAspThrSer 1164  
 3331 CATGGCCGAGCAGCAGCAGCTCCAGAGAGCTGTGATACAACTCC 3380  
 1164 YThrGlnHisProPheAlaProGluTrpProIleProGlyAlaSer 1181  
 3381 GACCACATCCCTTTGCTGAGACACCAATCCCTGAGCATCTGGA 3430  
 1181 eIleSerProTrpThrProGlyLysLeuProTrpGlyTrpThrGln 1197  
 3431 GCATCTCTCTACACCCCGGCGGCTGCTTGGGCTGAGATCTGAG 3480  
 1198 ProThrProValProGluAspLysGlyGlnProGlyLysLeuArg 1214  
 3481 CTTAGCGCAGTCCCTGAGGAGCAAAAGGCACTGAGAGAGACTGAG 3530  
 1214 sProGlyThrSerLeuProAlaAspLeuPro 1224  
 3531 TCCCGCAGCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3561  
 Seq\_name: gb\_pat:AX211284

seq\_documentation block:  
 LOCUS AX211284 3144 bp DNA linear PAT 06-SEP-2001  
 DEFINITION Sequence 1 from Patent WO0159133.  
 ACCESSION AX211284  
 VERSION AX211284.1 GI:15523693  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 3144)  
 AUTHORS Duecker K., Heutsch B., Hoheisel J. and Frohme M.  
 TITLE Larynx carcinoma-associated protein larcap-1  
 JOURNAL Patent: WO 0159133-A 1 (16-AUG-2001)  
 MERCK PATENT GmbH (DE)  
 FEATURES  
 source location/Qualifiers  
 1..3144  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..>3144  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC69262.1"  
 /db\_xref="GI:15523694"



/translation="MHASAAHSPVLAMLPWLAKGSOVILLIGPVCYLLEECPL  
LPQGRGVGGTLLGOMOLERYTGMLGNVAGGSIQSODEIYHDESLCVHINIALVRLIM  
VGYROSILIERGNSRSLRBOCRRMAHSQOROPSHAEHHDHVFLITRDPGSGTAP  
VTGMCHPLRSALNHEDESSAFYIAHETGVLHMEHDEGNGCADESLGSMALV  
OAAHRTHMSRCSKLEISRYLSDCLDDPFPAMPQPELPGIYNVMDQCREFPG  
SGYOTCLAFRTFEPCKQLMCSHPDNPFCKTKGKAPLDGECAPGKWCRCIMKSP  
EQTGQDGQWSMTKFGSGSRSGGVRSSRSCNNPSTAPGVRGLICLGMFEYQONS  
EECPGYEEDFRAOQCAKRSYVYHONAKHSMVYEPNDPAOKCELICQASADGVYPM  
NOVHDTGCSYRPSVCAFGECVPCDKEVSGSKADKCGVCGDNSHCRTVKT  
LGKSKAGALAKVQIPAGARHIOLEALEKSPHRIYVKNQVTSFLINKGKEATSRT  
FTMGLMEDVADEAKESLTSGLPEALAILALPTGEGRRSLAYKVIHEDLLPL  
IGSNVNLLEMDIYEWALIKSMAPCSKAGGGLQFTYCCRRRHHMVRHICDKKR  
PKPIRRRCNOHPCSOYVWTEEMWAGSRSRGLQYOTRGLOCLPLISNTHVPMKA  
CAGDRPARRRPLRVPDPAOMRLGAMSOKYLSTSCMDPLVLRMPESPINQELLAL  
VOPTVLCATGEGIOOROVYCRTNANSOKILSTSCMDPLVLRMPESPINQELLAL  
ADVWEICATGEGIOOROVYCRTNANSOKILSTSCMDPLVLRMPESPINQELLAL  
CVSCTKASASPENGPDPSTPLPSTPGSPDPDPAADAEPRGKPKGSDHOGHR  
ATDLPGLDITSPTQHPAPETPIPASMSISPTTGGGLPMGWTOTPTPVPEDEKGP  
GEDLRHPTSLPASPVT"

BASE COUNT 633 a 1000 c 957 g 554 t  
ORIGIN

alignment\_scores:  
Quality: 5160.50 Length: 970  
Ratio: 5.573 Gaps: 2  
Percent Similarity: 95.464 Percent Identity: 95.361

alignment\_block:

US-09-938-330-20 x AX211284 ..

Align seg 1/1 to: AX211284 from: 1 to: 3144

291 ValAspGluIleTyrHisAspGluSerLeuGluValHisIleAsnIleAl 307  
|||||  
238 GTAAGATGAGATTTCACAGATGAGTCCCTGGGGGTCAATATAATATGTC 287  
307 aLeuValArgLeuIleMetValGlyTyrArgGlnSerLeuSerLeuIleG 324  
288 CCTCCTCGCTGATCATGTTGGGTACGACGACGTCCTGAGCCTGATCG 337  
324 IuArgGlyAsnProSerArgSerLeuGluGlnValCysArgTTPAlaHis 340  
338 AGCGGGGAGACCCCTCACGACGCTGGAGGAGGTGCTCCTGGGCGCAC 387  
341 SerGlnGlnArgGlnAspProSerHisAlaGluHisAspHisValVa 357  
388 TCACAGACGCGCCAGGACCCAGCAGCTGAGCACCATGACACCGTTGT 437  
357 IPhleuThrArgGlnAspPheGlyProSerGlyTyrAlaProValThrG 374  
438 GTTCTCTACCCGCGAGGACTTTGGGCCCTCAGGGTATGACACCGTCACTG 487  
374 IYMetCysHisProLeuArgSerCysAlaLeuAsnHisGluAspGlyPhe 390  
488 GCATGTGTACACCCCTGAGGAGCTGTGCTCCTCAACCATGAGGATGGCTTC 537  
391 SerSerAlaPheValIleAlaHisGluThrGlyHisValLeuGluIleMetG 407  
538 TCCTCAGGCTTCGTGATAGCTCATGAGACCGGCGACCTGCTGGCATGGA 587  
407 uHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeuGlySerY 424  
588 GCATGAGGCTCAGGGAATGGCTGTGACATGAGACGACCTGGGACGAGG 637  
424 aIMetAlaProLeuValGlnAlaAlaPheHisArgPheHisIleSerArg 440  
638 TCATGGGGCCCTGTGTGTCAGGCTGCTTCACCGCTTCATTTGGTCCGCC 687  
441 CysSerIleuGluLeuSerArgTyrLeuProSerTyrAspCysLeu 457  
688 TGCAGCAAGCTGAGGCTCAGCGCTACCTCCTCTACGAGCTGCTCTCT 737

457 uAspAspProPheAspProAlaTyrProGlnProProGluLeuProGlyI 474  
|||||  
738 CGATGACCCCTTTGATCTGCTGGGCCAGCCCAAGAGCTGCTGGGA 787  
474 IeAsnTyrSerMetAspGluGlnCysArgPheAspPheGlySerGlyTyr 490  
788 TCACACTACTCAATGATGAGCAGTCCGCTTGGACTTTGGAGTGGCTCAC 837  
491 GlnThrCysLeuAlaPheArgThrPheGluProCysIleGlnLeuTyrPcy 507  
838 CAGACCTGCTTGGCATTTGAGACCTTTGAGCCCTTCAGAGCTGTGTTG 887  
507 sSerHisProAspAsnProTyrPheCysIleThrIleGlyIleProProL 524  
888 CAGCCATCTCTGACAAACCCGTACTTCTGCAGAACCAAGAGGAGGCCCCG 937  
524 euAspGlyThrGluCysAlaProGlyIleTyrPcyPheGlyHisCys 540  
938 TGGATGGAGCTAGTGTGACCCGCGCAAGTGTGCTTCAAGGTCCTGC 987  
541 IleTyrIleSerProGluGlnThrTyrGlnAspGlyGlyTyrPser 557  
988 ATCTGGAAGCTCGCCGAGACAGATATGCCAGATGAGGCTGAGACTC 1037  
557 rTTPhrIlePheGlySerCysSerArgSerCysGlyGlyValAlaArg 574  
1038 CTGGACCAAGTTGGGTGATGTCCTCGGCTCATGTGGGGGCGGGGTGCGAT 1087  
574 eArgSerArgSerCysAsnAsnProSerProAlaTyrGlyIleArgPro 590  
1088 CCGCGACGCGAGAGCTGCAACAAACCCCTCCAGCTTATGGAGGCGGCTG 1137  
591 CysLeuGlyProMetPheGluTyrGlnValCysAsnSerGluGlyCysPr 607  
1138 TGGTATGAGGCCCATGTTGAGTACAGAGTGTGCAACAGCGAGGAGTCCC 1187  
607 oGlyThrTyrGlnAspPheArgAlaGlnGlnCysAlaIleArgAsnSer 624  
1188 TGGGACCTACGAGGACTTCGCGGCCCAAGCATGTGCAACCGCAACTCCT 1237  
624 YrTyrValHisGlnAsnAlaIleHisSerTyrValProTyrGluProAsp 640  
1238 ACTATGTCACACAGATATGCCAAGCAGCTGGGTGCTTCAGAGCTGAC 1287  
641 AspAspAlaGlnIleCysGluLeuIleCysGlnSerAlaAspThrGlyAs 657  
1288 GATGACGCCCAAGAGTGTGAGCTGATCTGCCAGTGGCGGAGCACGGGGGA 1337  
657 pValValPheMetAsnGlnValValHisAspGlyThrArgCysSerTyr 674  
1338 CTTGGTGTTCATGAACCAAGGTGTTCAAGATGGAGACGCTGACGTACC 1387  
674 rGAspProTyrSerValCysAlaArgIleGluCysValProValGlyCys 690  
1388 GAGACCATACAGGCTGTGCGGGTGGCAGAGTGTGCTGCTGGCTGT 1437  
691 AspIleGluValGlySerMetIleAspAspIleCysGlyValIleCysG 707  
1438 GACAAAGAGGTGGGCTCCATGAAGGCGGATGACAAAGTGTGAGTGGCGG 1487  
707 YGlyAspAsnSerHisCysArgThrValIleGlyIleThrLeuGlyIleAla 724  
1488 GGTGTGACAACTCCACTGCGAGACTGTGAAGGGGAGAGCTGGGGAGAGCCT 1537  
724 eIleGlyGlnAlaGlyAlaLeuIleValGlnIleProAlaGlyAlaArg 740  
1538 CCAAGCAGCAGAGAGCTCTCAAGCTGTGCAAGTCCACAGAGGTGACAGG 1587  
741 HisIleGlnIleGluAlaLeuGluIleCysSerProHisArgSerValVal 757  
1588 CACATCCAGATTAAGGACTGAGGAGAGTCCCCCAGCCGATTTGGTGA 1637  
757 sAsnGlnValThrGlySerPheIleLeuAsnProIleGlyIleGluAlaT 774

```

1638 GAACAGAGTACCGGAGCTTCATCTCAACCCCAAGGCGAAGAACCA 1687
1688 hrSerArgThrPheThrAlaMetGlyLeuGluTrpGluAspAlaValGlu 790
1688 CAAAGCCGAGCTTCACCGCCATGGGCTGGAGTGGAGGATGGGGTGGAG 1737
1791 AspAlaLysGluSerLeuLysThrSerGlyProLeuProGluAlaIleAl 807
1738 GATGCCAAGAAAGCTCAAGAACCCAGCGGCGCCCTGCTCAAGCCATGTC 1787
807 alleuLeuAlaLeuProProThrGluGlyGlyLysProArgSerSerLeuAla 824
1788 CATCTGGCTCTCCCGCACTGAGGGGGGCCCCCGCAGCAGCTGGGCT 1837
824 yTrpLysTrpValIleHisGluAspLeuLeuProLeuIleGlySerAsnAsn 840
1838 ACAAGTACGTCATCCATGAGGACCTGGCTGCCCTTATCGGAGCAACAAT 1887
841 ValLeuLeuGluGluMetAspThrTrpGluTrpAlaLeuLysSerTrpAl 857
1888 GTGCTCCGAGAGAGATGACACCTATGAGTGGGCGCTCAAGAGCTGGGC 1937
857 aProCysSerLysAlaCysGlyGlyIleGlnPheThrLysTrpGlyC 874
1938 CCCCTGCAAGCAAGGCTGTGGAGGAGGATCCAGTTCACCAATAACGCT 1987
874 ySArGArGArGArAspHisSHMetValGlnAArgHisLeuCysAspHis 890
1988 GCGGCGGAGAGAGAGACCAACCATGATGAGCGACGACACCTGTGACAC 2037
891 LysLysArgProLysProIleArgArgArgCysAsnGlnHisProCysSe 907
2038 AAGAAAGAGGCCCAAGCCATCCGCGGCGTGCACCAAGCAAGCCGTGCTC 2087
907 rGlnProValTrpValIleThrGluGluTrpGlyAlaCysSerArgSerCysG 924
2088 TCAGCCTGTGGTGGAGGAGAGAGTGGGGTGCCTGCAGCCGAGCTGTG 2137
924 yLysLeuGlyValGlnThrArgGlyIleGlnCysLeuLeuProLeuSer 940
2138 GGAAGTGGGGGTGCAGACGCGGGGATACAGTCTGCTGCTCCCTCTCC 2187
941 AsnGlyThrHisLysValMetProAlaLysAlaCysAlaGlyAspArgPr 957
2188 AATGGAACCCACAGGTATGCCGCGCAAAAGCCTGCGCGGAGCGGCTC 2237
957 oGluAlaArgArgProCysLeuArgValProCysProAlaGlnTrpArgL 974
2238 TGGAGGCCAGCGGCGCTGTCTCGAGTGCCTGCGCCAGCCAGTGGAGGC 2287
974 euGlyAlaTrpSerGln..... 979
2288 TGGAGAGCTGTGCTCCGAATACTTCTGCTAGACACTAGCTGATGCCAGAC 2337
979 ..... 979
2338 CTGTGACTAAGATGAGGAACCTAGCATCAACAGACAGAGCTGATTTCT 2387
980 ..... CysSerAlaThrCysGlyGlyIle 988
2388 TGGCCCTGTGCACGCCACAGTCTTGTGCTGTGCACCTGTGGAGAGGCTA 2437
988 leGlnGlnArgGlnValValCysArgThrAsnAlaAsnSerLeuGlyHis 1004
2438 TCCAGCAGCGCGGAGGTGTGTGAGACCAACAGCCCAAGCCTCGGGCAT 2487
1005 CysGluGlyAspArgProAspThrValGlnValCysSerLeuProAlaCys 1021
2488 TGGAGAGGGAGATAGGCCAGACACTGTCCAGGTCTGCAGCCTGCGCCCTG 2537
1021 sGlyGlyAsnHisGlnAsnSerThrValArgAlaAspValTrpGluLeu 1038

```

```

2538 TGGAGGAATACACCAACTCCAGCGTGAGGCGCGATGTCTGGAACTTG 2587
1038 yThrProGluLysGlnTrpValProGlnSerGlyProLeuHisProIle 1054
2588 GAGCGCCAGAGGGGAGGTGGGTGCGCAATCTGAACCCCTACATCCATT 2637
1055 AsnLysIleSerSerMetCysAlaIleGluProCysTrpGlyAspArgSe 1071
2638 AACAAATATCATCA.....ACGAGCCCTGCAGGAGACAGGCTC 2678
1071 rValPheCysGlnMetGluValLeuAspArgTrpCysSerIleProGlyT 1088
2679 TGTCTTCTCCAGATGAGAGTCTGATGCTGATCTGATCAAGAGGCTCGGG 2728
1088 yHisArgLeuCysCysValSerCysIleLysLysAlaSerGlyProAsn 1104
2729 ACACCGGCTGTCTGTGTCTGTCTGATCAAGAGGCTCGGGGCCCAAC 2778
1105 ProGlyProAspProGlyProThrSerLeuProProPheSerThrProG 1121
2779 CTGGCCCAAGACCCCTGGCCCAACCTCACTGCCCCCTTCTCCACTCTGG 2828
1121 ySerProLeuProGlyProGlnAspProAlaAspAlaIleGluProProG 1138
2829 AAGCCCTTACAGAGACCCAGAGACCTGACATGCTGAGAGGCTCTGG 2878
1138 LysProThrArgLysSerGluAspHisGlnHisGlyValArgAlaThrGln 1154
2879 GAAGGCCAAGGGATCAGAGACCATCAGCATGAGGCGGAGCAACAGCTC 2928
1155 ProGlyAlaLeuAspThrSerSerProGlyThrGlnHisProPheAlaPr 1171
2929 CAGAGAGCTGTGATACAGCTCCCAAGGAGACCCAGACATCCCTTTCCCC 2978
1171 oGluThrProIleProGlyAlaSerTrpSerIleSerProThrThrProG 1188
2979 TGAAGACCAATCCCTGGAGCATCTGGAGCATCTCCCTACACACCCCG 3028
1188 yGlyLeuProTrpTrpGlyTrpThrGlnThrProThrProValProGluAsp 1204
3029 GGGGGCTGCTTGGGGCTGAGCTCAGACACCTAGCGCAGTCCCTGAGAG 3078
1205 LysGlyLysProGlyGluAspLeuArgHisProGlyThrSerLeuProAl 1221
3079 AAGGGCAACCTGAGAGACCTGAGACATCCCGGACACGCTCCCTGCG 3128
1221 aAspLeuPro 1224
3129 TGCCTCCCG 3138
seq_name: gb_pat:E58661
seq_documentation_block:
LOCUS E58661 3618 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel metallic protease.
ACCESSION E58661
VERSION E58661.1 GI:18629883
KEYWORDS JP 2001017183-A/9.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3618)
AUTHORS Yamaji,N., Nishimura,K. and Sasamata,M.
TITLES Novel metallic protease
JOURNAL Patent: JP 2001017183-A 9-23-JAN-2001;
YAMANOUCHI PHARMACEUT CO LTD
COMMENT OS Homo sapiens (human)
PN JP 2001017183-A/9
PD 23-JAN-2001
PF 09-JUL-1999 JP 1999196584
PR NOBORU YAMAJI, KOICHI NISHIMURA, MIHO SASAMATA

```

PC C12N15/09, C07K16/40, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC  
 C12N9/50, C1201/37,  
 PC C12N15/00, C12N5/00  
 CC  
 FH Key Location/Qualifiers  
 FT source 1..3618  
 FT /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /organism="Homo sapiens (human)"

FEATURES  
 source 1..3618  
 location/Qualifiers  
 BASE COUNT 1035 a 802 c 880 g 901 t  
 ORIGIN

ment\_scores:  
 Quality: 3624.00 Length: 1198  
 Ratio: 4.081 Gaps: 17  
 Percent Similarity: 74.124 Percent Identity: 55.593

alignment block:

US-09-938-330-20 x E58661 ..

Align seg 1/1 to: E58661 from: 1 to: 3618

```

37 AspTYGlyValThrValProCysSerThrAspPheArgGlyArgPheLe 53
||||| .....| | | | | | | | | | | | | | | | | | | | |
121 GAGTATGAGCTGTGACTCCAGTCCAGCAATATCAGAAAGAGAGCTATCT 170
53 uSerHisValValSerGlyProAlaAlaAlaSerAlaGlySerMetValV 70
||||| .....| | | | | | | | | | | | | | | | | | | | |
171 CTCCTATACCTCTTCT..... 186
70 AlaSPThrProProThrLeuProAlaArgHisSerSerHisLeuArg..... 84
187 .....ValAlaArgSerProLeuHisProGlyThrLeuTrpPr 207
85 .....| | | | | | | | | | | | | | | | | | | | |
208 GCGAGGAGCGTGTCTTCCACCT..... 231
98 oGlyArgValGlyArgHisSerLeuTrpPheAsnValThrValPheGlyL 115
232 .....GAGCAGTTGTTCTTTAACATCATCAGCGCATTTGGAA 265
115 ySGluLeuHisLeuArgLeuArgProAsnArgArgLeuValValProGly 131
||||| .....| | | | | | | | | | | | | | | | | | | | |
266 AAGATTTCATCTGCGACTAAGCCCAACACTCAACTAGTAGCTCCG 315
132 SerSerValGluTrpGlnGlu..... 138
316 GCTGTGTGGAGTGCATGAGACATCTCTGTGCTGGCAATATTAACCGA 365
139 .....AspPheArgGluLeuPheA 145
366 TCCCATTAACACCATCAACACGAGAACTGCTACGTTATAGAAATCCGGA 415
145 rGlnProLeuArgGlnGlyValValThrGlyValValThrGlyMet 161
416 CAGAGCCTTTGGACACTAAGTGTATGTGTGATGACATCGAGCAATT 465
162 ProGlyAlaAlaValAlaIleSerAsnGlyAspGlyLeuAlaGlyLeu 178
||||| .....| | | | | | | | | | | | | | | | | | | | |
466 CAGAGAACTCTGTGCTTGCATGAGCACTGATGATGCTGCTGCGAATGAT 515
178 eArgThrAspSerThrAspPhePheLeuGluProLeuGluArgGlyGlnG 195
||||| .....| | | | | | | | | | | | | | | | | | | | |
516 AAAAAGATGATATCAAGAGATTTTCATGTAGAACCTTGAAAGAGTAAC 565
195 InGluysGluAlaSerGlyArgThrHisValValValArgArgGluAla 211
||||| .....| | | | | | | | | | | | | | | | | | | | |
566 AGATGGAGGAAGAAAGAAAGATTCATGTGCTTACCAAGAGATCAGCT 615

```

```

212 ValGlnGlnGluTrpAlaGluProAspGlyAspLeuHis.....As 225
||||| .....| | | | | | | | | | | | | | | | | | | | |
616 GTAGAACAGAGCTCCCATAGACATGTCCAAAGAGCTTCCATCAGAGATC 665
225 nGluAlaPheGlyLeuGlyAspLeuProAsnLeuGlnGlyLeuValGly 242
||||| .....| | | | | | | | | | | | | | | | | | | | |
666 GAGCCTGGAGAGGCTTGTATGATGATGCTGTATGCAAAATCCAC 715
242 sPGLuLeuGlnAspThrGluArgLysArgArgHisAlaLysProGlySer 258
||||| .....| | | | | | | | | | | | | | | | | | | | |
716 AGCAGCTGATGATAAACAATGAGAGCCCGAGACGCGGAGAAACGAT 765
259 TyrSerIleGluValLeuLeuValAlaAspAspSerValValArgPhe 275
||||| .....| | | | | | | | | | | | | | | | | | | | |
766 TACAATATCAGAGTACTGTGCTGAGTGTGATGATGATGCTGCTTCCA 815
275 sGlyLysGluHisValGlnAsnTrpValLeuThrLeuMetAsnIleVal 292
||||| .....| | | | | | | | | | | | | | | | | | | | |
816 TGCAAAAGAGCAGTCCAAAACTACCTCTGACCTTAATGAAATCTGA 865
292 sPGLuIleTyrHisAspGluSerLeuGlyValHisIleAsnIleAlaLeu 308
||||| .....| | | | | | | | | | | | | | | | | | | | |
866 ATGAATTTACCATGATGATGCTCCGAGTGCATATTAATGTGCTCG 915
309 ValArgLeuIleMetValGlyTrpArgGlnSerLeuSerLeuIleGluAr 325
||||| .....| | | | | | | | | | | | | | | | | | | | |
916 GTGCCATGATGATATGCTGATGATGCAAAAGTCCATCAGCTCATAGAAG 965
325 gGlyAsnProSerArgSerLeuGlnValCysArgTrpAlaHisSerG 342
||||| .....| | | | | | | | | | | | | | | | | | | | |
966 GCGAAACCATTCAGAAAGCTTGGAAATGTGTGCTGGGCGTCCCA 1015
342 InGluArgGlnAspProSerHisAlaGluHisHisAspHisValValPhe 358
||||| .....| | | | | | | | | | | | | | | | | | | | |
1016 AGCAAAAGATGTGATCTCAACCATCTGTACCAACATGCAATTTT 1065
359 LeuThrArgGlnAspPheGlyProSer.....GlyTrpAlaProVa 372
||||| .....| | | | | | | | | | | | | | | | | | | | |
1066 TTAAACAGGCAAGACTTTGGACCTGTGGAATGCAAGATATGCTCAGT 1115
372 ThrGlyMetCysHisProLeuArgSerCysAlaLeuAsnHisGluAsp 389
||||| .....| | | | | | | | | | | | | | | | | | | | |
1116 CACCGGCGATGTGTATCATCAGAGAGAATGTGTACCTTAATCATGAGAG 1165
389 lPheSerSerAlaPheValIleAlaHisGluThrGlyHisValLeuGly 405
||||| .....| | | | | | | | | | | | | | | | | | | | |
1166 GTTTTTCATCTGCTTTGTGTAGTACCGAAGCGGCCATGTGTGGGA 1215
406 MetGluHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeu 422
||||| .....| | | | | | | | | | | | | | | | | | | | |
1216 ATGAGCATGATGAGCAAGCAAGGATGTGTGATGAGACTCTCTGTGG 1265
422 ySerValMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTrp 439
1266 AAGTGTATGCTCTCTGTGTACAGCAGCATTCATTCGTATCCATCTGT 1315
439 eArgCysSerLysLeuGluLeuSerArgTrpLeuProSerTrpAspCys 455
||||| .....| | | | | | | | | | | | | | | | | | | | |
1316 CCGATTCAGAGGTGCAAGAACTGAAAGATATATCATTCCTATGACTGT 1365
456 LeuLeuAspAspProPheAspProAlaTrpProGlnProProGluLeuPr 472
||||| .....| | | | | | | | | | | | | | | | | | | | |
1366 CTCTGTATGACCTTTTGTATCATGATTTGGCTTAACCTCCAGAACTTGC 1415
472 oGlyIleAsnTyrSerMetAspGluGlnCysArgPheAspPheGlySerG 489
||||| .....| | | | | | | | | | | | | | | | | | | | |
1416 TGCATTCATTAATTCATGATGAGCAATGCTGTTTGTATTTGTGTGTG 1465
489 lYTrpGlnThrCysLeuAlaPheArgThrPheGluProCysLysGlnLeu 505
||||| .....| | | | | | | | | | | | | | | | | | | | |
1466 GCTATTAAGTGTGACCGGCTTCCAGAACTTTGACCCATGTAAGAGCTG 1515
506 TrpCysSerHisProAspAsnProTyrPheCysLysThrLysGlyLys 522

```

```

|||||
1516 TGGTGTAGCATCCTGATTAATCCCTTTGTAAGACTAAAGGAC 1565
522 oProLeuAspGlyThrGluCysAlaProGlyLysTrpCysPheLysGly 539
|||||
1566 TCCACTTGATGGAGCTGATGTGCTGCTGGAATAATGCTGTATAAAGGCTC 1615
539 IScySIIeTrpLysSerProGluInThrTrpGlyGlnAspGlyGlyTrp 555
|||||
1616 AATGCAATGTGAAGATGCTAATCG...CAAAACCAAGATGGCAATGG 1662
556 SerSerTrpThrLysPheGlySerCysSerArgSerCysGlyGlyGly 572
|||||
1663 GGGTCATGAGCTAAATTTGGCTCTCTCTCGGACATGTGGAAGCTGTGT 1712
572 IArgSerArgSerArgSerCysAsnAsnProSerProAlaTrpGlyGly 589
|||||
1713 TCGTTTCAGAACCGCAGCTGCATTAATCCCATGCCATCATGCTGTGTC 1762
589 rGpTrpCysLeuGlyProMetPheGluTrpGlnValCysAsnSerGlu 605
|||||
1763 AGGATTTGCTCGTGTAAATTTTGAATACAGCTTTGTAACACGAAAGAA 1812
606 CysProGlyThrTrpGluAspPheArgAlaGlnGlnCysAlaLysArg 622
|||||
1813 TGCACAAAACACTTTGAAGACTTCAGAGCACAGCAGTGTCAGCAGGAAA 1862
622 nSerTrpTrpValHisGlnAsnAlaLysHisSerTrpValProTyrGlu 639
|||||
1863 CTCCCACTTGAATATCCAGAAATACCAACACACTGGTTGCCATATGAA 1912
639 rObAspAspAlaGlnLysCysGluLeuLysCysGlnSerAlaAspThr 655
|||||
1913 ATCTGACCCCAAGAAAGATGCCACCTTTACTGTCACTCCAGAGACT 1962
656 GlyAspValValPheMetAsnGlnValValHisAspGlyThrArgCys 672
|||||
1963 GAGAGATGTTGCTTACATGAACAACCTGTCATGATGAGAACGACTGTC 2012
672 rTyTrpArgProTyrSerValCysAlaArgGlyGluCysValProVal 689
|||||
2013 TTACAAACATCATATACATATGTGTGCGAGAGAGTGTGTAAAGTGG 2062
689 LysAspLysGluValGlySerMetLysAlaAspAspLysCysGlyVal 705
|||||
2063 GCTGTGATTAAGAAATGTGTTCTATATAGTGTGAGATTAAGTGTGTC 2112
706 CysGlyGlyAspAsnSerHisCysArgThrValLysGlyThrLeuGly 722
|||||
2113 TGTGAGAGGATTAATCCCACTGCCGAACCGTGAAGGGACATTTAAC 2162
722 sAlaSerLysGlnAlaGlyAlaLeuLysLeuValGlnLeuProAlaGly 739
|||||
2163 AACTCCCAAGAAAGCTTGGGTACTTAAGATGTTGTTATTAACCCCTG 2212
739 IArgHisIleGlnIleGlnAlaLeuGlnLysSerProHisArgSerVal 755
|||||
2213 CTAGACATGTGTTATCCAGAAAGACGAGCTTCCCTCATATCTTCTCT 2262
756 ValLysAsnGlnValThrGlySerPheIleLeuAsnProLysGlyLys 772
|||||
2263 ATTAAAGAACCGCTAACAGGCTATATATTAAATGGCAAGGGAGAGA 2312
772 uAlaTrpSerArgTrpPheThrAlaMetGlyLeuGlnTrpGluAspAla 789
|||||
2313 AGCCAAAGTGGGACCTTCATAGATCTTGGTGTGAGTGGGATTTAAC 2362
789 aGlnuAspAlaLysGlnSerLeuLysTrpSerGlyProLeuProGlnAla 805
|||||
2363 TTGAAGATGACATTAAGTCTTCAACCGATGAGACCTTTACATGATCT 2412
806 ILeuAlaIleLeuAlaLeuProThrGlnGlyGlyProArgSerLeu 822
|||||
2413 GTTATGTTTGAATATA...CTCCAAAGAAATGATACCCGCTCTAGCCT 2459
822 uAlaTrpTrpTrpValIleHisGluAspLeuLeuProLeuIleGlySer 839
|||||
2460 GACATATAAGATGATCATTCATGAAAGACTCTGTACTCATTCACAAACA 2509
839 sAsnValLeuLeuGlnLysMetAspThrTrpGluTrpAlaLeuLysSer 855
|||||
2510 ACAATGTATCCAGAGAAATATGATCTTTGATGTGGCTTGAAGAGC 2559
856 TrpAlaProCysSerLysAlaCysGlyGlyIleGlnPheThrLysTrp 872
|||||
2560 TGTCTCACTGTTCCAAACCTGTGTGTGAAGTTTCCAGTACACTAAATA 2609
872 rGlyCysArgArgArgArgAspHisMetValGlnArgHisLysCys 889
|||||
2610 TGGATCCCTTAGAAAGATTAATAATGCTGCATGACGACTTGTGTG 2659
889 sPHisLysLysArgProLysProIleArgArgArgCysAsnGlnHisPro 905
|||||
2660 AGGCCAAACAAAAGCCGAACCTATTAGACGAATGTGCATATTTCAAG 2709
906 CysSerGlnProValTrpValThrGlnGlnTrpGlyAlaCysSerArg 922
|||||
2710 TGTACACATCCACTGTGGGTACAGAAAGATGGGAACCTGCACCAAA 2759
922 rCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysLeuLeuPro 939
|||||
2760 CTGTGGAAGTTCTGCGTATACAGCTTCGACGTGACGCTCTTCAAGCA 2809
939 eUeSerAsnGlyThrHisLysValMetProAlaLysAlaCysAlaLys 955
|||||
2810 TCCCTGATGACCAACCAACCGCTCTGTGCACAGCAAAATCTGTGGTG 2859
956 ArgProGlnAlaArgArgProCysLysLeuArgValProCysProAlaGln 972
|||||
2860 CGTCCCGAGAGCCCGCGCCCTTACACAGATGCCCTGCGCCGACAGTG 2909
972 rArgLeuGlyAlaTrpSerGlnCysSerAlaThrCysGlyGlyGlyIle 989
|||||
2910 GAACAACGACCTCGAGTGAAGTGTTCAGTGCACCTCGGTGAAGAACCG 2959
989 IlnAlaArgGlnValCysArgThrAsnAlaAsnSerLeuGlnHisCys 1005
|||||
2960 AGGTGAGGCGGCTCTGTGCGAGGGCTGGG...GACCACTGT 2997
1006 GlnGlyAspArgProAspThrValGlnValCysSerLeuProAlaCys 1022
|||||
2998 GATGTGAAGAAGCTGAGTGTGTCAGAGCCTGTCACTGCTCTGTGA 3047
1022 yGlyAsnHisGlnAsnSerThrValArgAlaAspValTrpGluLeuGly 1039
|||||
3048 TGAT... 3051
1039 hrProGlnGlyGlnTrpValProGlnSerGlyProLeuHisProIleAsn 1055
3051 ..... 3051
1056 LysIleSerSerMetCysAlaAlaGluProCysThrGlyAspArgSer 1072
|||||
3052 .....GAAACATGTTGGAGACAAAGTCAT 3077
1072 LpHeCysGlnMetGlnValLeuAspArgTrpCysSerIleProGlyTrp 1089
|||||
3078 ATTCTCAATGGAAGTGTGGCAGCATCTGCTCCATACAGGTTTAA 3127
1089 sArgLeuCysCysValSerCysIleLysLysAlaSerGly... 1102
|||||
3128 ACAAGTATGTGTGAGTCTGTGACGACGACGACGTACACCCGCCACCA 3177
1103 .....ProAsnPr 1105
|||||
3178 CCATACCTTGTAGAAAGTGTGAACCTCATGATGATGTCTATTAACC 3227

```

```

1105 oclyProaspProgly.....ProthSerleuProrophe. 1117
1118 .....SerThProgly.....SerProleu 1124
1125 ProglyProglinspProAlaAlaIaIuProProglylySProth 1141
1141 rglYsergluAspHsglnHsglAArgAlaThrglnleuProglyAlaI 1158
1158 euAspThr.....SerSerProglyThrgln 1166
3419 AGACTGTGAGACTGTGACCGTACCATCTCCACCACCACCAAG 3462
seq_name: gb_pr:AB002364

seq_documentation block:
LOCUS AB002364 5774 bp mRNA linear PRI 06-OCT-2001
DEFINITION Human mRNA for KIAA0366 gene, partial cds.
ACCESSION AB002364
VERSION AB002364.1 GI:2224672
KEYWORDS KIAA0366.
SOURCE Homo sapiens male brain cDNA to mRNA, clone_1lb:pluescriptII SK
plus clone:HH0124.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 5774)
AUTHORS Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
VII. The complete sequences of 100 new cDNA clones from brain which
can code for large proteins in vitro
JOURNAL DNA Res. 4 (2), 141-150 (1997)
MEDLINE 97349987
REFERENCE 2 (bases 1 to 5774)
AUTHORS Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
TITLE Direct Submission
JOURNAL Submitted (28-Mar-1997) Osamu Ohara, Kazusa DNA Research Institute;
1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdna@info.kazusa.or.jp, Tel:+81-438-52-3913)
FEATURES
location/Qualifiers
source 1..5774
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HH0124"
/sex="male"
/tissue_type="brain"
/clone_lib="pluescriptII SK plus"
1..3607
/gene="KIAA0366"
<1..3607
/gene="KIAA0366"
/codon_start=2
/db_xref="GI:2224672"
/translation="STMIAALVETVSADGAGNEENVQIDLPKRYREVELVTPV
STNLEGRSLHTLSASHKRSARSDVSNBPDLFPNTAFGDFHLKRPNTOLVAGA
VWEHETSLVPGNTIDP INNHQPSATYRIRTEPIOTNCAVGDIVDIPGVALSN
CDGLAGMIKSDNEYETIEPLERKGEERGHVYVYKRSAYVADIDMSKDFHONY
LLEGLDGLGYVNIHOOLNTEBRBRHAGENYNIIEVLGVDSYVFRHGEHYON
LTLEMTVNETYHDESGVHINVLVYMLTGLAKSTSLIERGNPSRSLENCNMAQO
QORSDLNHSEHDAIFLTRODFGPAMQGVAVTGMCHPVSCTLNHEDGSSAFVY
AHETGHVLMGHDQGRGDETMASVMAPIVOAFAFHRHMGSCGGLKRYISYD
CLDDPDHDMPKLPELPGINYSMEDECRDFDVGAKMCTAFPTDPCGLMCSHPDN
PYFCKTKKGPPLDTECAAGKWCYKKGCMKNNNOCKOGGNNMSWTKPSCSRCTGTG

```

```

VERTRQCNPNMP INGGDPCGVNEFYOLCNTEECKHFEDEFRQAQCCORNSHPEYON
TKHMLPJEHPKPKRCHLYCQSKETGVADYKMLQVHDGTCYKDPYSICVRECYK
VCGDKIEGSKNEDKCGVCGGNSHCRTKGTFTPKRILYLMKMFDPGARHVILLO
EDASPILIAIKNAQNGHYILNGKGEERKSRFTDGLVEMOYNIEDOIESLHTGPRH
DPYIVLITROENDTRBSLTKYKTHIEDSPITNSNNVIOBELDTEBALMSVOSVSRP
CGGEPYITTCGRKRSKDNKMWIRSFCEANKKPKPIRMNCNIOECTHPLMAEWEHCT
KTGSSGYOLRTVRLDPLDGTNRSVHSKYCOMGRPSRRPCNRNVPDPAMKTPWMS
ECSVTCGEETEVROYLCRAGDHDEKPEPVACOLPPCNDPCLGKSIICOMEVLA
RYSCLPGNKLCCESCSKRSSTLPPVYLLEAEETHDVIYSNPSDLPSILVMTSLVPY
HSETPAKKMSLSSISVSGNAYAAFRNSRPGDANKIORSAAQOAGSKTYVRLVPS
PPTKRYHSSASOMAAFPFASDSIGASSQARTSKKQKIIIDNRPRSTLTER"
BASE COUNT 1766 a 1129 c 1258 g 1621 t
ORIGIN
alignment_scores:
Quality: 3615.50 Length: 1274
Ratio: 3.956 Gaps: 22
Percent Similarity: 71.743 Percent Identity: 53.373
alignment_block:
US-09-938-330-20 x AB002364 ..
Align seg 1/1 to: AB002364 from: 1 to: 5774
37 AspyrGlyValThrValProGlySerThrAspPheArgGlyArgPheLe 53
110 GAGTATGAGCTGTGTGACTCCAGTCCAGCAAAATCTGAAAGAGCGTATCT 159
53 userHisValSerGlyProAlaAlaIaSerAlaGlySerMetValV 70
160 CTCCCATACCTCTTCT..... 175
70 alaThrProProThrLeuProArgHisSerHisLeuArg..... 84
176 .....ValAlaArgSerProLeuHisProGlyGlyThrLeuTrp 98
85 .....ValAlaArgSerProLeuHisProGlyGlyThrLeuTrp 98
197 GCGAGGACGCTGTCTTCCAAACCT..... 220
98 oGlyArgValGlyArgHisSerLeuTrpPheAsnValThrValPheGlyL 115
221 .....GAGCAGTTGTTCTTTAACATCAGCGCATTTGGAA 254
115 ysgIuLeuHisLeuArgLeuArgProAsnArgArgLeuValValProgly 131
255 AAGATTTTCATCTGCGACTAAAGCCCAACACTCACTAGTACTCTCTGGG 304
132 SerSerValGluTrpGlnGlu..... 138
305 GCTTGTGTGAGTGGCATGAGCATCTCTGTCGTCGGAATAATACCA 354
139 .....AspPheArgGluLeuPheA 145
355 TCCCATTAACACCATCAACACAGAGAGTCTAGTATAGATACCGGAAA 404
145 rglInProLeuArgGlnGluGlyValTyrThrGlyGlyValThrGlyMet 161
405 CAGAGCGCTTTGACACTGACTGCTTATGTTGGTACATCGTGACATT 454
162 ProGlyAlaAlaValAlaIaIeSerAsnGlySAspGlyLeuAlaGlyLeu 178
455 CCAGGAACCTCTGTTCATCAGCAACTGTATGTTGCTGCGTGGAAATAT 504
178 eAgtThrAspSerThrAspPhePheIleGluProLeuGlnArgGlyGln 195
505 AAAAAGTGAATATGAAGACTATTTCATTCAGACCTTGAAGAGAGTAAAC 554
195 InGluysGlnAlaSerGlyArgGlnHisValValTyrArgArgGlnAla 211
555 AGATGGAGGAGAAAGAAAGAGATTTCATGTTGCTTACAGAGATCAGCT 604

```

```

212 ValGlnGlnIuTrpAlaGluProAspGlyAspLeuHis.....As 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
605 GTAGAAAGAGGCTCCCATAGACATGTCCAAAGACTTCCACATACAGAGAGTC 654
225 nGluAlaPheGlyLeuGlyAspLeuProAsnLeuLeuGlyLeuValGlyA 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
655 GGACCTTGGAAAGGCTTGTATGTATAGTACTGTATTATGGCAACATCCACC 704
242 sPGLuLeuGlyAspThrGluArgLysArgArgHisAlaLysProGlySer 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
705 AGCAGCTGATGAAACAAATAGACAGCGCGACAGACGGGGGAAAAAGCAT 754
259 TyrSerIleGluValLeuLeuValValAspAspSerValValArgPheH 275
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
755 TACAAATACGAGTACTGTGGAGTGTAGTACTGTGTGTCCGTTCCCA 804
275 sGlyLysGluHisValGlnAsnTyrValLeuThrLeuMetAsnIleValA 292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
805 TGGCAAAAGACGCTCCAAAGACTTCTGACCTTATGCAACATTTGTCA 854
292 sPGLuIleTyrHisAspGlySerLeuGlyValHisIleAsnIleAlaLeu 308
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
855 ATGAATTTTCCATGATGAGTCCCTCGAGTGCATATAAATGTGTCCTG 904
309 ValArgLeuIleMetValGlyTyrArgGlnSerLeuSerLeuIleGluA 325
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
905 GTGGCATGTATGATCTGGATATGCAAGTCCATCAGCTCATAGAAAG 954
325 gGlyAspProSerArgSerLeuGluGlnValCysArgTyrAlaHisSerG 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
955 GGGAAACCCATCCAGAACCTTGGAGAAATGTGTGTGCGTGGGCGTCCAC 1004
342 IngInArgGlnAspProSerHisAlaGluHisHisAspHisValValPhe 358
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1005 AGCAAAAGATCTGATCTGACCACTGTGACCACTGACCACTGCAATTTT 1054
359 LeuThrArgGlnAspPheGlyProSer.....GlyTyrAlaProVa 372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1055 TTAAACAGGACAGACTTTGGACCTGTGTAATGCAAGATATGCTCCAGT 1104
372 ThrGlyMetCysHisProLeuArgSerCysAlaLeuAsnHisGluAspG 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1105 CACCGGATGTGTATCCAGTGAAGAGTTGTACCTGATCATGAGAGTGT 1154
389 LysPheSerSerAlaPheValIleAlaHisGluThrGlyHisValLeuGly 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1155 GTTTTTCATCTGCTTTGTAGAGCCATGAAAGCGGCATGTGTGGGA 1204
406 MetGluHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeuG 422
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1205 ATGGAGCATGATGGACAAGGACAGTGTGTGTATGAGACTGCTATGGG 1254
422 ySerValMetAlaPheValGlnAlaAlaPheHisArgPheHisTrpS 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1255 AAGTGTCTATGGCTCCCTGGTACAGACAGCATTCATGTTACACATGGT 1304
439 eArgCysSerLysLeuGluLeuSerArgTyrLeuProSerTyrAspCys 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1305 CCCGATGCAAGTGTCAAGACAGAAAGATATATCCATTCCTATGACGT 1354
455 LeuLeuAspAspProPheAspProAlaTrpProGlnProProGluLeuP 472
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1355 CTCCTTGATGACCTTTTGTATGATGAGCTTAACATCCAGAACTTCC 1404
472 oGlyLysAsnTyrSerMetAspGluGlnCysArgPheAspPheGlySerG 489
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1405 TGGAAATCAATTAATCTATGATGACCAATGTCTTTGTATTTTGTGTTC 1454
489 LysTyrGlnThrCysLeuAlaPheArgThrPheGluProCysLysGlnLeu 505
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1455 GCTATTAATATGTGCACCGGCTTCCGACACTTGTGACCATGTAAACAGCTG 1504
506 TrpCysSerHisProAspAsnProTyrPheCysLysThrLysLysGlyP 522

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1505 TGTGTACACCTCCGTAATATCCCTACTTTGTAAAGACTAAAGGAGCC 1554
522 oProLeuAspGlyThrGluCysAlaProGlyLysTrpCysPheLysGlyH 539
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1555 TCCACTTGATGGAGCTGTAATGTGTGTGCGAAATGGCTATTAAGGCTC 1604
539 LscGlyIleThrLysSerProGluGlnThrTyrGlyGlnAspGlyGlyTyr 555
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1605 ATTGATGTGGAAAGATCTTATTCAG...CAAAACAGATGGCAATTTGG 1651
556 SerSerTrpThrLysPheGlySerCysSerArgSerCysGlyGlyValA 572
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1652 GGGTCATGAGCAATAATTGGCTCCTGTCGAGACATGTGCAACGTGCT 1701
572 LArgSerArgSerArgSerCysAsnAsnProSerProAlaTyrGlyValA 589
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1702 TCGTTTCAGAAACCGCCAGTGCATTAATCCATGCGCATCAATGGTGTG 1751
589 rGProCysLeuGlyProMetPheGluTyrGlnValCysAsnSerGluGlu 605
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1752 AGGATGTCTCTGTGTTAATTTTGTAGTACCACTTTGTAAACAGAAAGA 1801
606 CysProGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaLysArgAs 622
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1802 TGCAGAAAACACTTTGAGGACTTCAGAGCACAGAGTGTCAAGCAGAAA 1851
622 nSerTyrTyrValHisGlnAsnAlaLysHisSerTyrValProTyrGluP 639
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1852 CTCCCACTTTGAATCCAGAAATACCAACACACAGCTGTGTTCATTTGAC 1901
639 roAspAspAspAlaGlnLysCysGluLeuLeuLeuGlnSerAlaAspThr 655
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1902 ATTCGTGACCCCAAGAAAGATGCCACTTACTGTGACGTCAAGAGACT 1951
656 GlyAspValValPheMetAsnGlnValValHisAspGlyThrArgCysSe 672
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1952 GGAGATGTGCTTACATGAACACAGCTGTGATGATGAGAACACACGTGTC 2001
672 rTyrArgAspProTyrSerValCysAlaArgGlyGluCysValProValG 689
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2002 TTACAAAGATCCATATAGCATATGTGTGCGAGAGAGTGTGAAGTGG 2051
689 LysCysLysGluValAlGlySerMetLysAlaAspLysCysGlyValA 705
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2052 GCTGTATTAAGAAATGTTGTTAATAGTGTGAGATGAAGTGGTGTGTC 2101
706 CysGlyGlyAspAsnSerHisCysArgThrValLysGlyThrLeuGlyL 722
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2102 TGTGAGAGAGATATTTCCACTGCCAGACCGTGAAGGGAGACATTACCG 2151
722 sAlaSerLysGlnAlaGlyAlaLeuLysLeuValGlnIleProAlaGlyA 739
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2152 AACTCCAGAGACCTTGGTACCTTAAGATGTTGTATATACCCCTGGGGG 2201
739 LArgHisIleGlnIleGlnAlaLeuGlnLysSerProHisArgSerVal 755
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2202 CTAGACATGTGTTAATCCAGAAAGAGGCTTCTCTCATATTCTTGTCT 2251
756 ValLysAsnGlnValThrGlySerPheIleLeuAsnProLysGlyLysG 772
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2252 ATTAAAGACAGGCTACAGGCTTATATTATTTAAATGGCAAGGAGAGA 2301
772 uAlaThrSerArgThrPheThrAlaMetGlyLeuGluTyrPGLuAspAla 789
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2302 AGCCAAATGCGGACCTTCATAGATCTTGTGTGAGTGGAGATTATTAACA 2351
789 aGluAspAlaLysGlnSerLeuLysHisSerGlyProLeuProGluAla 805
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2352 TTGAAGATGACATTTGAAGAGTCTTACACCGATGACCTTTTACATGATCT 2401
806 IleAlaIleLeuAlaLeuProThrGluGlyGlyProArgSerSerLe 822
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

2402 GTTATTGTTTGGATTATA...CCTCAAGAAAATGATACCCGCTTAGCCT 2448  
 822 uAlaTYrLYsTYrVAlIleHISgluSpleuLeuProLeuIIegLYSeR 839  
 2449 GACATATTAATACATCATATGAAGATCTGTACTACATCAATCAACACACA 2498  
 839 snAsnValIleuLeuGluMetAspThrTYrGluTYrAlaLeuLYsSer 855  
 2499 ACAATGTCATCCAGAGAAATTAAGTACTTTGAGGGGCTTTGAGAGGC 2548  
 856 TrpAlaProCYsSerLYsAlaCYsGlyGlyIleGlnPheThrLYsTY 872  
 2549 TGGTCTCAGAGTTCCAAACCCCTGTGTGGAGGTTTCCAGTACATAAATA 2598  
 872 rGlyCYsArgArgArgAspHisMetValGlnArgHisLeuCYsA 889  
 2599 TGGATCCCTAGGAAAGATGATATAAATGATGTCACAGCTCTGTGTG 2648  
 889 sPHISLYsArgProLYsProIleArgArgCYsAsnGlnHisPro 905  
 2649 AGGCCCAACAAAAGCCGAACCTATTAGACGAATGCAATATTCAAGAG 2698  
 906 CYsSerGlnProValTrpValThrGluGluTrpGlyAlaCYsSerArgSe 922  
 2699 TGTACATTCACCTCTGGGTAGCAGAGAATGGCAACACTGCACCAAAAC 2748  
 922 rCYsGlyLYsLeuGlyValGlnThrArgGlyIleGlnCYsLeuLeuProL 939  
 2749 CTGTGAAATCTCTGGCTATACACTTCGCACACTGTACGCTGCTTCAACCC 2798  
 939 euSerAsnGlyThrHisLYsValMetProAlaLYsAlaCYsAlaGlyAsp 955  
 2799 TCCCTGATGGCACCAACCGCTCTGTGCACAGCAAACTACTGCATGGGTGAC 2848  
 956 ArgProGluAlaArgArgProCYsLeuArgValProCYsProAlaGlnTr 972  
 2849 CGTCCCGAAGAGCCGCCGCCCTGTACAGAGCTGCCCTGCTGCACAGTG 2898  
 972 pArgLeuGlyAlaTrpSerGlnCYsSerAlaThrCYsGlyGluGlyIleG 989  
 2899 GAAACAGAGACCCCTGAGTGAAGTGTTCAGTACCTGGCGGTAAAGAACGG 2948  
 989 InGlnArgGlnValValCYsArgThrAsnAlaAsnSerLeuGlyHisCYs 1005  
 2949 AGGTAGGAGAGTCTCTGTCAGGGCTGTGC.....GACCACCTGT 2986  
 1006 GluGlyAspArgProAspThrValGlnValCYsSerLeuProAlaCYsG 1022  
 2987 GATGGTGAAGAACCTGAGTCCGTACAGACCTGTCAACTGCTCTTGTAA 3036  
 1022 yGlyAsnHisGlnAsnSerThrValArgAlaAspValTrpGluLeuGlyT 1039  
 3037 TGTAT..... 3040  
 1039 hrProGluGlyGlnTrpValProGlnSerGlyProLeuHisProIleasn 1055  
 3040 ..... 3040  
 1056 LysIleSerSerMetCYsAlaIleGluProCYsThrGlyAspArgSerVa 1072  
 3041 .....GAACCATGTTGGGAGAACAGATCCAT 3066  
 1072 IPhcCYsGlnMetGluValLeuAspArgTYrCYsSerIleProGlyTYrH 1089  
 3067 ATTCTGTCAATGAAGTGTGGCAGATACTGCTCATACCAAGGTATATA 3116  
 1089 IAsArgLeuCYsCYsValSerCYsIleLYsLYsAlaSerGly..... 1102  
 3117 ACAAGTTATGTTGTAGTCTCGACAGCAAGCGAGTACACCCGTCACACA 3166  
 1103 .....ProAsnPr 1105  
 3167 CCATACCTTTAGAAAGCTGCTGAACATCATGATGATCTATCTCTAACCC 3216

1105 ogLYsProAspProGly.....ProHisSerLeuProProphe. 1117  
 3217 TAGTACGATCCCTAGATCTCTAGTATGCTCATATCTTTGGTCTCTATTC 3266  
 1118 .....SerThrProGly.....SerProLeu 1124  
 3267 ATTCAAGAGACCCCTGCMAAGAAATGCTTTTGTAGTACATCTCTTCACTG 3316  
 1125 ProGlyProGlnAspProAlaAspAlaIleGluProProGlyLYsProTh 1141  
 3317 GAGAGTCCAAATGCATATCT...GCTTTCAGGCCCAAAACAGTAAACCTGA 3363  
 1141 rGlySerGluAspHisGlnHisGlyArgAlaThrGlnLeuProGlyAlaL 1158  
 3364 TGGTCTCAATTTTACGCCAG.....AGAGTGTCTCAGCAACAGAGGAAGTA 3407  
 1158 euAspThr.....SerSerProGlyTYrGlnHisPro 1168  
 3408 AGACTGTGAGACTGTGACCGTACATCTCTCCACCAACAGAGAGGTC 3457  
 1169 .....PheAlaProG 1172  
 3458 CACCTCAGTTCACTTCAACAAATGCTGCTGCTTCTCTTTCAGACCCAG 3507  
 1172 uhrProIleProGlyAlaSerTrpSerIleSerProThrThrProGlyG 1189  
 3508 TATTCATTAATGAGTGTCTTCTTCAGCAGCAAGAACTCAAGAAAGATGAA 3557  
 1189 LYeu.....ProTrpGlyTrpThrGlnThrProThr... 1199  
 3558 AGATCATTTGCACACAGACTCCG.....ACAGATCATTCACCTTA 3598  
 1200 .....ProValPro.....GluAspLYsGlyGlnProG 1209  
 3599 GAAAGATGAGAAATGAGAACCAAAAGCTGACAAACAGAGAGAAACCTGG 3648  
 1209 yGluAspLeuArgHisProGly 1216  
 3649 AACACCTCTCTCTTCCATGTT 3670

seq\_name: gb.com:BTCPINPcn  
 seq\_documentation\_block: 4580 bp mRNA linear MAM 01-MAR-1997  
 LOCUS BTCPINPcn  
 DEFINITION B.taurus mRNA for procollagen I N-proteinase.  
 ACCESSION X96389  
 VERSION X96389.1 GI:1865715  
 KEYWORDS PC I-NP gene; procollagen I N-proteinase.  
 SOURCE cow.  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 4580)  
 Colige A., Nussgens B. V. and Lapiere C.M.  
 Cloning of the cDNA of the bovine procollagen I N-proteinase  
 2 (bases 1 to 4580)  
 Colige A.C.  
 Direct Submission  
 Submitted (29-FEB-1996) A.C. Colige, University of Liege,  
 Laboratory of Connective Tissue Biology, Tour de Pathologie, B23/3  
 - CHU Sart-Tilman, B-4000 Sart-Tilman, BELGIUM  
 location/Qualifiers  
 1. 4580  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /tissue\_type="skin"  
 /dev\_stage="foetal"  
 138..3755  
 /gene="PC I-NP"  
 138..3755  
 CDS







```

429 ValGlnAlaIaIaPheHisArgPheHisIleTyrSerArgCysSerIleLeuG1 445
1434 GTGCAGGACAGCTTCCATCGCTCCAGCTGTCGCCGTGAGCAGCAGCAGCA 1483
445 uIeuSerArgTyrLeuProSerTyrAspCysIleuLeuAspPheProPhea 462
1484 GCTTACGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1533
462 sPProIaTyrProGlnProGlnIleuLeuProGlnIleuLeuSerIleTyr 478
1534 CCACAGACAGCTGCGGCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1583
479 AspGlnGlnCysArgPheAspPheGlySerGlyTyrGlnIleuLeuAla 495
1584 AACGACAGCTGCGCTTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1633
495 aPheArgTyrPheGlnProCysGlnIleuTyrCysSerHisProAspA 512
1634 GTTCCGACCTTGCAGCCGTGCAACAGCTGCTGCTGCTGCTGCTGCTGCT 1683
512 sNProTyrPheCysIleuTyrIleuLeuProIleuAspGlyTyrGln 528
1684 ACCCTGCTTTCGCAAGACAAAGAGGCGCCACCTGCTGCTGCTGCTGCTG 1733
529 CysAlaProGlyIleuTyrCysPheIleuLeuIleuLeuIleuLeuIleu 545
1734 TGTGCGCCGCGCAGCAGCTGCTTAAAGACACTGCTGCTGCTGCTGCTG 1783
545 oGlnGlnIleuTyrGlnIleuLeuIleuLeuIleuLeuIleuLeuIleu 562
1784 TACAC...ATTTCAAACAGAGAGGCACTGCTGCTGCTGCTGCTGCTG 1830
562 IySerCysSerArgSerCysGlyGlyValArgSerArgSerArgSer 578
1831 GCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1880
579 CysAsnAsnProSerProAlaTyrGlyValArgProCysLeuGlyProme 595
1881 TCCGACAAACCCACACCCAGCAATGGGCGCGCACATGCTGCGGCTGCG 1930
595 tPheGlnIleuValCysAsnSerGlnIleuCysProGlyTyrIleuAla 612
1931 CTACGATTTCCAGCTCTCTCAACTGCGAGACTGCTGCTGCTGCTGCTG 1980
612 sPHeArgIleuGlnIleuCysAlaIleuAsnSerTyrTyrValIleuGln 628
1981 ACTTCCGCGAGAGAGAGTGCAGGAGTGCAGTCTGCTGCTGCTGCTGCT 2030
629 AsnAlaIleuHisSerTyrValProTyrGlnProAspAspAlaGlnIleu 645
2031 GACGCTCAACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2080
645 sCysGlnIleuIleuCysGlnSerAlaAspThrGlnIleuValIleuPhe 662
2081 GTGTCATCTCTACTGTAGTGCAGAGACCCGGGAGTGTGTCTCATGCA 2130
662 sNglIleuValIleuHisAspGlyTyrArgCysSerTyrArgAspProTyr 678
2131 AGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2180
679 ValCysAlaArgGlyIleuCysValProValGlyCysAspIleuValG 695
2181 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2230
695 ySerMetIleuAlaAspAspIleuCysGlyValCysGlyGlnIleuAsp 712
2231 CTCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2280
712 IyCysArgTyrValIleuGlyTyrIleuGlyIleuValIleuValIleu 728
2281 ACTGCAAGGTGCTGCAAGGCAAGTCTGCTGCTGCTGCTGCTGCTGCTG 2330

729 AlaIleuIleuValGlnIleuProAlaGlnIleuAlaArgHisIleuGln 745
2331 TACATACAGAGTGTGTGATATCCGCGAGGAGCAGCAGCAGCAGCAGC 2380
745 uAlaIleuGlnIleuSerProHisArgSerValIleuIleuValIleu 762
2381 GGAAGCAGACACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2430
762 IySerPheIleuLeuAsnProIleuGlyIleu...GlnAlaThrSerArg 777
2431 GCAAGTTCAATTTTAAATGAGAGAAATGACGTGATCCCACTCCAAAG 2480
778 PheThrAlaMetGlyIleuGlnIleuProGlnIleuValIleuAlaIleu 794
2481 TTCATGCGCATGGGCTGAGTGGAGTACCGGATGAGGAGGAGGAGGAG 2530
794 uSerIleuIleuTyrSerGlyProIleuProGlnIleuAlaIleuAla 811
2531 GACGCTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2580
811 euProProThrGlnIleuGlyIleuProArgSerSerIleuAlaTyrIleu 827
2581 TCCCA...GAGGGGAGAGCGCCGATCTCATCTGACCTTACAAATGAT 2624
828 ILeuIleuAlaIleuProIleuIleuIleuIleuIleuIleuIleu 844
2625 ATCCATGAGAGACTCCCTCAAT...GTGATGACAAACAGCCTGAGAGA 2671
844 uGlnIleuAspThrTyrGlnIleuProAlaIleuIleuIleuIleuIleu 861
2672 CGACTCTGAGGCTATGATGAGGCGCTGCAAGAGAGTGTGCTGCTGCT 2721
861 yAlaIleuGlyIleuGlyIleuIleuIleuIleuIleuIleuIleu 877
2722 ACCCTGCTGCTGAGGAGTCCCAATTCACCAAGTATGCTGCTGCTGCT 2771
878 ArgAspHisIleuMetValGlnArgHisIleuCysAspHisIleuAsp 894
2772 CTGAGACCAACAGATGTGACACCGAGGCTTGTGCTGCTGCTGCTGCT 2821
894 oLysProIleuArgArgArgCysAsnGlnHisIleuProCysSerGlnPro 911
2822 CAAGCCATCCCGCGGAGCTGCAACCCAGAGAGTGTGCTGCTGCTGCT 2871
911 rPValIleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleu 927
2872 GGGTACGCGGTGAGTGGAGCGGTGAGCGGAGCGGAGCTGCTGCTGCT 2921
928 ValGlnIleuArgIleuGlnIleuCysIleuLeuProIleuSerGlnIleu 944
2922 ATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2971
944 sLysValMetProAlaIleuAlaCysAlaGlnIleuAspArgProGlnIleu 961
2972 CCGCTCGGTGACACCAAGCAGCAGTGCATGACGCTGACCCGAGGCG 3021
961 rGProCysIleuArgValProCysProAlaGlnIleuProGlnIleuAla 977
3022 GGGCTGCAACCGGAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3071
978 SerGlnCysSerAlaIleuIleuIleuIleuIleuIleuIleuIleu 994
3072 TCCAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3121
994 IyCysArgTyrAsnAlaAsnSerIleuGlnIleuIleuIleuIleuIleu 1011
3122 CTGCTGCAACGCGAGCAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCT 3171
1011 sPThrValGlnIleuIleuIleuIleuIleuIleuIleuIleuIleu 1027
3172 AGAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3221
1028 SerThrValArgAlaAspValIleuGlnIleuIleuIleuIleuIleu 1044

```



175 GCGGTGCCCGCTGCCACTGACGCCAGGCCCGCTGGTGTCCACGTGG 224  
 57 ISeGlyProAlaIaIaSerAlaGlySerMetValValAspThrProp 74  
 225 CTCGGCA...GCTACGTCCAGAGAGGGGTACAGAGCCCGAGGGCGGCC 271  
 74 roThrLeuProArgHisSerSerHisLeuArgValAlaArgSerPro 90  
 272 CG.....GTCGGAGCCCGAGAC 288  
 91 HisProGlyGlyThrLeuTrpProGlyArgValGlyArgHisSerLeu 107  
 289 TTCCTCCGGAGGCAAC.....GAGGAGAGCCCGGACATCAC...CTCTT 329  
 107 rPheAsnValThrValPheGlyLysGluLeuHisLeuArgLeuArgPro 124  
 330 CTACAATGTCACGCTCTTGTGGCCAGACCTGCACCTGGCGTGGCCCA 379  
 124 snArgArgLeuValValProGlySerSerValGluTrpGlnLysPhe 140  
 380 ACGCCGCGCTCGTGGCCCGGCGCATTATGAGATGGCAGGGCGAGAG 429  
 141 ArgGluLeuPheArgGlnProLeuArgGlnGlyValTyrThrGly 157  
 430 GGCACCAACCGCGCTGGAGCCCTGCGGAGCTGTCTTACGTGGAGA 479  
 157 yValThrGlyMetProGlyAla...AlaValAlaIleSerAsnGlyAsp 173  
 480 CGTGGCGCGCTATGCGGAGGCTCTCTGTGGCGCTGCACAACTCCGATG 529  
 173 LysLeuAlaGlyLeuIleArgThrAspSerThrAspPheIleGluPro 189  
 530 GCGTGGCTGTGTGATCCGATGAGGAGGAGGAGGAGGAGGAGGAGG 579  
 190 LeuGluArgGlyGlnGlnGlyLysGluAla...SerGlyArgThrHisVa 205  
 580 TTGGAGAAAGGGCTGGCGGCGAGAGGCTGAGAGGCTGAGCGGCTG 629  
 205 LValTyrArgArgGluAlaValGlnGlnGluTrpAlaGluProAspGly 221  
 630 GGGATATCCCGCGCACCCACCTCCCTCGGCGGCGCACAGGCC 679  
 222 ..AspLeuHisAsnGluAlaPheGlyLeuGlyAspLeuProAsnLeu 237  
 680 TTGACACAGAGGCTCCCTCGGACAGCTGAGAGCTTCAGCGCGCTG 729  
 238 GlyLeuValGlyAspGlnLeuGlyAspThrGluArgLys...ArgArgHis 253  
 730 GCGTCTCTAGAGAGCAGCGCCACAGCTCGAGGGGAGGAGGAGGAGCA 779  
 253 sAlaLysProGlySerTyrSerIleGluValLeuLeuValAlaAspAs 270  
 780 TCTGTGACAGCATGTAGTACTACAACTGAGGTCTCTGTGGCTGATGACT 829  
 270 eValValArgPheHisGlyLysGluHisValGlnAsnTyrValLeuThr 286  
 830 CTGTGGTCAGTTCACGAGGAGGAGGAGCTACAGAAAGTACTCTTGACA 879  
 287 LeuMetAsnIleValAspGluIleTyrHisAspGluSerLeuGlyValHis 303  
 880 CTCATGAACATTGTCAATGAATCTACCATGACGAGTCTTGGGTGCCA 929  
 303 sIleAsnIleAlaLeuValArgLeuIleMetValGlyTyrArgGlnSer 320  
 930 CATCAAGCTGTCTCTGGTGGCATCATCTCTCTAGCTATGGAAGATCCA 979  
 320 euSerLeuIleGluArgGlyAsnProSerArgSerLeuGlnValLys 336  
 980 TGAGCTCATGAGATCGGAAACCTCTCAGAGCCTGGAATGTGTGC 1029  
 337 ArgTrpAlaHisSerGlnGlnArgGlnAspProSerHisAlaGluHisHis 353  
 1030 CGGTGGCGCTTACTCTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1079  
 353 sAspHisValValPheLeuThrArgGlnAspPheGlyProSer..... 367  
 1080 CGATCAGCATCTCTCTCTACAGCGGAGGAGCTTTGGCGCTTCCGGATTC 1129  
 368 ..GlyTyrAlaProValThrGlyMetCysHisProLeuArgSerCysAla 383  
 1130 AAGGTATGCTCTGTACACCGCATGTGCTGCTGCTGCTGCTGCTGCTG 1179  
 384 LeuAsnHisGluAspGlyPheSerSerAlaPheValIleAlaHisGluThr 400  
 1180 CTGACCATATAGAGACGGCTTCTCTACGCTTTGTGGTGGCCATGAGAC 1229  
 400 rGlyHisValLeuGlyMetGluHisAspGlyGlnGlyAsnGlyCysAla 417  
 1230 TGGCAGCTGTGGCATGAGAGACAGCAGGAGGAGGAGGAGGAGGAGGAG 1279  
 417 spGluThrSerLeuGlySerValMetAlaProLeuValGlnAlaAlaPhe 433  
 1280 ACAGGATGGCGCTGGGACATCATGTGGCGCCCTGTGTGAGCGCGCTTC 1329  
 434 HisArgPheHisTrpSerArgCysSerLysLeuGluLeuSerArgTyrLe 450  
 1330 CACCGCTTCCAGTGTCCGCGTGCAGGCGAGCAGAGAGCTGAGCCGCTACT 1379  
 450 uproSerTyrAspCysLeuLeuAspAspProPheAspProAlaTrpPro 467  
 1380 GCATCTCTATGACTGTCTGTGATGAGACCCCTCCGACAGCATGGCGG 1429  
 467 InProGluLeuProGlyIleAsnTyrSerMetAspGluGlnCysArg 483  
 1430 CGGTGGCCCAAGCTCCCGGAGCTGCACTACTCTCATTAAGAGCAATAGCCG 1479  
 484 PheAspPheGlySerGlyTyrGlnThrCysLeuAlaPheArgThrPheG 500  
 1480 TTGTACTTGGCGCTGGCTATCATGTGACAGCGGCTTCGAGACTTTGA 1529  
 500 uproCysLysGlnLeuTrpCysSerHisProAspAsnProTyrPheCysL 517  
 1530 CCCTGTCAAGCAGCTGTGGTGGAGCCATCTGACACACCTTACTTTTGA 1579  
 517 ystThrLysLysGlyProProLeuAspGlyThrGluCysAlaProGlyLys 533  
 1580 AGACAAAGAAAGGGGCGCCCTTGAGCGGAGCTATGTGTGACCTGGGAG 1629  
 534 TrpCysPheLysGlyHisCysIleTrpLysSerProGluGlnThrTyrG 550  
 1630 CATTTTAAAGGACACTGCATCTGGCTGACACCTGAC...ATCTCAA 1676  
 550 yGlnAspGlyGlyTrpSerSerTrpThrLysPheGlySerCysSerArg 567  
 1677 ACGGAGCGGAGCTGGGGCGCTGTGGAGTCCGCTTGGCTCTGCTCACAGTA 1726  
 567 eCysGlyGlyGlyValArgSerArgSerArgSerCysAsnAsnProSer 583  
 1727 CCGTGGACAGCGGCGGTGAAGTTGAGGAGCCCGCAGTGTGACAAACCCAC 1776  
 584 ProAlaTyrGlyArgProCysLeuGlyProMetPheGluThrGlnVa 600  
 1777 CCGGAGCAAGGGGGCGGCACTGTCTGGGCTTCCCTTACACATTCACAGCT 1826  
 600 LysAsnSerGlyLysProGlyThrTyrGluAspPheArgAlaGln 617  
 1827 CTGACGCGGAGGAGCTGCGGCTCCGCTGCTGCTGCTGCTGCTGCTGCTG 1876  
 617 LncysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHisSer 633  
 1877 AGTGGCGCAATGGAGCTTACTTCCAGACGCGGAGGAGGAGGAGGAGGAG 1926  
 634 TrpValProTyrGluProAspAspAlaGlnLysCysGluLeuLeu 650  
 1927 TGGCTGGCCCAAGAGCAGCGGAGTGCACAGAGAGATGTCACCTGTACTG 1976







2976 ACTCTGTCGC.....TGCC 2989

1036 IuleuGlyThr.....ProGluGlyGlnThrProGlnSerGly... 1049  
|||||  
|||||  
2990 CCTCTGGCGACACTGCGCCCTGAAGGCTCAGGACGGGCTCTCCACGCAC 3039  
1050 .....ProLeuHisProLeuAsnLysI 1057  
3040 GAGCTCTTACAGACGCTACTTCATTCGCCGACACCTGGCCCCACGCC 3089  
1057 eSerSerMetCysAlaIaGlnProCysThrGlyAspArgSerValPheC 1074  
|||  
::: |||  
3090 TTCACCCGCTCATCACCCACGACGACGACCATGGGCAAGCATTTAGG 3139  
1074 ysgInMetGluValLeuAspArgTyrCysSerIleProGly..... 1087  
::: |||  
3140 AGAGAGCTCCAGAGCTGGAC.....CTGGCGGGGCCCTGT 3177  
1088 .....TyrHisArgLeuGlyCysValSerCysIleLysLysAl 1100  
|||  
::: |||  
3178 GTGGAGCACTTCTACTAGACTACATTCATTCATTTCCACAGAGATCT 3227  
1100 aSer...GlyProAsnProGlyProAsp..... 1108  
|||  
|||  
3228 GTCTCTACGGGCCCTGTGAGGACCCGATCTAGACCTGGCGGACAGGG 3277  
1109 .....ProGlyProThrSerLeuProProPheSerThrProGlySer 1122  
|||  
|||  
3278 ACCGACACCCCAACACACACCATCTCTGTGGCCCTCCAGGGTAGC 3327  
1123 ProLeuProGlyProGlnAspProAlaAspAlaIaGlu.....Pr 1136  
|||  
|||  
3328 CCGTGTCCCTCCACAGAGCCTCTCCACACCAGAGAGAGGGGTACTGGG 3377  
1136 oProGlyLysProThrGlySerGluAspHisGlnHisGlyArg..... 1150  
|||  
|||  
3378 ACCTGTGTCCTCCGAGAC...CCTGGCCTAGCCAGCGCGCGCTGCCAC 3424  
1151 .....AlaThrGlnLeuProGly...AlaLeuAspThrSerSerPro 1163  
::: |||  
3425 CCCCACCTCTCAGAGACACCCCTGGGAACCTTTGATCATTTCCGCT 3474  
1164 GlyThrGlnHisProPhe...AlaProGluThrProIleProGlyAlaSe 1179  
|||  
|||  
3475 GAGGAGACACCCCATAGGGGCCCAAGATCTTGGGCTCCCGAGCTGTC 3524  
1179 rTTPser.....IleSerProThrThrPro.... 1187  
|||  
|||  
3525 CTGGGCCAGGGTTTCCACTGATGGCTTCGACAGACACCTGCCACCCTGAGA 3574  
1188 .....GlyGlyLeuProThr 1192  
3575 GCCAAATGATTTCCAGTTGGCAGAGACAGCCAGCAGCAGCTGCCCT 3624  
1193 GlyTTPThrGlnThrProThrProValProGlnAspLysGlyGlnProGl 1209  
|||  
|||  
3625 CCATGGCGGACAGGACCAATGAGGTTTTCAGAGATGATGAGAACCC... 3672  
1209 yGluAspLeuArgHisProGlyThrSerLeuProAlaAspLeuProGlyA 1226  
|||  
|||  
3673 .....AAGGCGCGGAGGACCC...CACCTGCCCGCA 3703  
1226 rgProProGlu.....ProCysHisProThrGlyThr 1236  
|||  
|||  
3704 GACCCAGCTCCAGCGTCCCTTTTGTCCCTGTGGCAGC 3744

seq\_name: gb\_pat:AX128433

seq\_documentation block:

LOCUS AX128433 5353 bp DNA linear PAT 15-MAY-2001

DEFINITION Sequence 2 from Patent WO0131034.

ACCESSION AX128433

VERSION AX128433.1 GI:14134924

[illegible]



```

169 TCGTGCACCCGGTTGAGTCGACGCGGGGGCTCTCTCTCTCTAC... 215
57 ValSerGlyProAlaAlaIleSerAlaGlySerMetValAlaSerThr 73
215 ..... 215
73 oProThrLeuProAlaHisSerSerHisLeuArgValAlaArgSerPro 90
215 ..... 215
90 euHisProGlyGlyThrLeuThrProGlyArgValAlaArgHisSerLeu 106
216 ..... GAGCTGTGGCCCGCCGACATCGCCAAAGCGGAGTGA 251
107 ..... TyrPheAsnValThrValPheGly 115
252 TCTGTGGCCGAGACGCGCCCTCTACAGCTCAATACCGCGGGCG 301
115 sGluLeuHisLeuArgLeuArgProAsnArgArgLeuValAlaProGly 132
302 CGAGCTGGCTTCAACCTGACCGCCCAATCAGCACCTCTGCGCGGCT 351
132 erSerValGluThrProGluAspPheArgGluLeuPheArgGlnProLeu 148
352 TTGTGAGCGAGACGCGGGCGG... GCGGCGCTGGGCGCGCGCACATC 398
149 ArgGlnGlu..... CysValTyrThrGlyValThrGly.... 160
399 CGGGCCACACCCCGCCCTGCTGCTGCGAGGTGCGAGACCTGA 448
161 MetProGlyAlaAlaValAlaIleSerAsnCysAspGlyLeuAlaGly 177
449 GCTCGAGGTGGCTGCGCGCATTCAGCTGCGACGCGCTGAAAGTG 498
177 euLeuArgThrAspSerThrAspPhePheLeuGluProLeuGlnArgGly 193
499 TGTTCACACTCTCAACGAGGACTCTCATTTGAGCCCTTGAC... AGT 545
194 GlnGlnGluGlyGluAlaSerGlyArgThrHisValAlaTyrArgArgGly 210
546 GCGCGCGCGCGCTGGCCACGCGCCACCGCTGCTGTGTGTACAAACGCTCA 595
210 uAlaValGlnGlnGluThrAlaGluProAspGlyAspLeuHisAsnGlu 227
596 GCGCGCGGAGAGGCTGCGACACAGCGGGGTGATTCAGTGTCCAAACACT 645
227 lApheGlyLeuGlyAspLeuProAsnLeuLeuGlyLeuValGlyAspGln 243
646 GT... GAGTGTGCAAGTGTACCCAGAGCTG... 671
244 LeuGlyAspThrGluArgLysArg..... 251
672 ..... GAGTCTCGACGGAGCGTTGGAGACGCGGCAAGTGGCGGG 715
252 ..... ArgHisAlaLysProGlySerTyrSer..... I 261
716 GCCACGCGTGAAGCGCTCTACACACGCGGTGCGTCAAGCAAGAGAGTGG 765
261 leGluValLeuLeuValAlaAspAspSerValAlaArgPheHisGly 277
766 TGGAGACCTGTGTAGTGTGCTGATGCCAAATGTGTGAGTACCAAGAGAG 815
278 GluHisValGlnAsnTyrValLeuThrLeuMetAsnLeuValAspGlu 294
816 CCGCAGCTTGAGAGCTATGTCTGACCATCATGATCATGCTGGCGCT 865
294 eTyrHisAspGluSerLeuGlyValHisIleAsnIleAlaLeuValArg 311
866 GTTTCATGACCCGACGATGGAACCCCATTCACATCACCATTGCGCGC 915
311 euIleMetVal... GlyTyrArgGlnSerLeuSerLeuIleGluArgGly 326
916 TGGTCTGTGGAAGATGAGAGAGGAGCACTTAAAGATCAGCAGCATGCA 965

327 AsnProSerArgSerLeuGlnValLysArgTrpAlaHisSerGlnGly 343
966 GAC..... AACCCCTGAGAGCTTCTGCAAGTGGCAAGAAAGCATCA 1009
343 nArgGlnAspProSerHisAlaGluHisHisAspHisValAlaPheLeu 360
1010 CATGAAGGGGAGTGGCCATCCCTGCACATGACATGCTCATCTGTCTCA 1059
360 hrArgGlnAspPheGlyProSer..... Gly 368
1060 CCAGAAAGAGCCTGTGTGACGACATGAACCGCCCTGTGAGACCTTGGGA 1109
369 TyrAlaProValThrGlyMetCysHisProLeuArgSerCysAlaLeuAs 385
1110 CTGTCCCATGTGGCGGACATGTGCGACCGCAGCCAGCTGAGCATCA 1159
385 nHisGluAspGlyPheSerSerAlaPheValIleAlaHisGluThrGly 402
1160 CGAGGACACGGGCTGCCCTGCTGCTTACTGTACCCAGCAGAGCTGGAGC 1209
402 iValLeuGlyMetGluHisAspGlyGlnGlyAsnGlyCysAlaAspGlu 418
1210 ACAGTTTGGCATTCAGATGACGGAAGCGGCAATGACTGTAGCCCTT 1259
419 ThrSerLeuGlySerValMetAlaProLeuValGlnAlaAlaPheHisAr 435
1260 GGAAGAACGACTTTCATCATGTCTCCACAGCTCTGACGAGCGCTTC 1309
435 gPheHisTrpSerArgCysSerIleLeuLeuSerArgTyrLeuPro 451
1310 CCTCACCTGCTCCGCTGCGACCGCATGTATATCCACAGATTCCTTGACC 1359
452 .. SerTyrAspCysLeuLeuAspAspProPheAspProAla..... 464
1360 GTGGGTGGGCGCTGTGCTGAGAGACCT.... CTGCCAAGGACATT 1403
465 ..... TrpProGlnProProGluLeuProGlyTyrIleAsnTyrSerMetAs 479
1404 ATGCACCTTCCCTCGGTGCCA..... CTTGGGCTCTTATCATGATTAAG 1447
479 pGluGlnCysArgPheAspPheGlySerGlyTyrGlnThrCysLeuAla 496
1448 CACACGAGTGGCGCTCCAGTACGG... ..GCT 1476
496 heArgThrPhe..... GluProCysLysGlnLeuTyrCys 507
1477 ACTGTGCTCTTCTGAGAGACATGATATGTCTGCCACACACTGTGTGC 1526
508 SerHisProAspAsnProTyrPheCysLysThrLysGlyProProLe 524
1527 TCTGTGGGACCAAC... ..TGTACTCCAGCTGTGATGCACCT 1567
524 uAspGlyThrGluCysAlaProGlyLysTyrCysPheLysGlyHisCys 540
1568 GGAGCGACCCGCTGTGGGAGATAGTGTGTCTCAGTGGGAGAGTGC 1617
541 ..... lEtrPlysSerProGluGlnThrTyrGlyGlnAspGlyGlyTyr 555
1618 TACCCTGTGGCTTCCGCGCCGAGCGCTG... ..GATGTGTGCTGG 1658
556 SerSerTrpThrLysPheGlySerCysSerArgSerCysGlyGlyGly 572
1659 TCTGTGGAGCGCTGTGCTCATCTGCTACGAGAGCTGTGCTGGAGCT 1708
572 lArgSerArgSerArgSerCysAsnAsnProSerProAlaTyrGlyGly 589
1709 ACAGAGCGCGGAGCGGACGTGACAGCAGCTTACGCCCAATACAAAGGCA 1758
589 rGProCysLeuGlyProMetPheGluTyrGlnValCysAsnSerGluGlu 605
1759 GATACCTGTGTGGTGAAGCGGCTTCCGCTTCCGCTTCCGCTTCCGCT 1808

```



2667 CCCTGCCCCGCGAC... TGGTGGCAGGAGTATGGCAGACTGTGCTCCAG 27133  
921 gSerCysGlyuLysLeuGluValGlnThrArgGlyIleGlnCysLeuLeuP 938  
2714 CTCCTGCGGGGCTGTGGGGCTCTCCCGCCGCGGGCGTGTGTCATGCCGA 2765  
938 rOleuSerAsnGlyThrHisLysValMetProAlaLysAlaLysAlaGly 954  
2764 GCGTGGGGCTGGATGACACAGAGGCGCTGGAGCCACCGCTGTAAAC 28133  
955 . . . AsparProGluAlaAspArgProCysLeuArg . . . ValProCysPr 969  
2814 CTCGCCGGCCCCCTACTAGAACCCCTTGGAACCCGCAATGTACCTGTCC 2865  
969 oAlaGluThrArgLeuGluValAlaTrpSerGlnCysSerAlaThrCysGly 986  
2864 GGCACACCTGGGCTGTGGGAACTGGTCAAGTCTACAGACATGTGGG 29133  
986 IuGlyIleGlnGlnArgGlnValValCysArgThrAsnAlaMetSerLeu 10022  
2914 AGGCACATCAGAGCGCGGAATGTCTCTGC . . . ACCAATGACACCGGTCTC 2960  
1003 GlyHisCysGluGlyAspArgProAspThrValGlnValCysSerLeuP 1019  
2961 CCTGTGACAGAGCGCCAGCGACGCCACCGCAAGTCACTGCTCTCTCC 3010  
1019 oAlaCysGlyGlyAsnHisGlnMetSerThrValArgAlaAspValITrp 1036  
3011 ACTGTGTCCG . . . . . TGGC 3024  
1036 IuLeuGlyThr . . . . . ProGluGlyGlnTrpValProGlnSerGly . . 1049  
3025 CCTGGGCAACATGGGCCCTGAAGGCTCAGGAGGGGTCTCCAGGCAC 3074  
1050 . . . . . ProLeuHisProIleAsnLysIle 1057  
3075 GAGCTCTTACAGAGGCTGACTTCATCTCCGCAACCCAGGCCACGCCC 3124  
1057 eSerSerMetCysAlaAlaGluProCysThrGlyAspArgSerValPheC 1074  
3125 TTCACCGCGCTCATCCGCCAGCCAGGCAACCATGGGGCAACCCATTGAG 3174  
1074 ySerGlnMetGluValLeuAspArgTyrCysSerIleProGly . . . . . 1087  
3175 AGGAGGCTCCAGAGCTGGAC . . . . . CTGCGCGGGCGCCGTGTTT 3212  
1088 . . . . . TyrHisArgLeuCysCysValSerCysIleLysLysAl 1100  
3213 GTGGAGCAGCTCTACTACAGACTACATTCATTCATATTTTCCACGGAGATCT 3262  
1100 aSer . . . GlyProAsnProGlyProAsp . . . . . 1108  
3263 GTCTCTACGGGCGCTCTGAGAGACCCGATCTAGACGTGGCGGGACAGGG 3312  
1109 . . . . . ProGlyProThrSerLeuProProPheSerThrProGlySer 1122  
3313 ACCGACACCCCCACACACACACGCTCTGTGCGCCCTCCAGGGGTAGC 3362  
1123 ProLeuProGlyProGlnAspProAlaAspAlaLacIu . . . . . Pr 1136  
3363 CCGTGTCTGTCACAGAGAGCTCTGCACCAAGAGAGAGGGGTACTGGG 3412  
1136 oProGlyLysProThrGlySerGlnAspHisGlnHisGlyArg . . . . . 1150  
3413 ACCTGTATGCCGAGC . . . . . CCTGGGCTTAGCCAGCGCGCGCTCCCCAC 3459  
1151 . . . . . AlaThrGlnLeuProGly . . . . . AlaLeuAspThrSerSerPro 1163  
3460 CCCCAACCTTCAAGACAGACCCCTGGAAACCTTTGATACATTTCTCTGCT 3509  
1164 GlyThrGlnHisProPhe . . . . . AlaProGluThrProIleProGlyLase 1179

```
3510 GAGGAAGACACCCCATAGGGGCCCAAGATCTGGGCTCCCCAGCCTGTC 3559
1179 fThpSer.....IleSerProThrThrPro.... 1187
      |||||
3560 CTGGCCCAAGGGTTTCCACTGATGGCTGCAGACACTGCCCACCCCTGAGA 3609
1188 .....GlyGlyLeuProTrp 1192
      |||||
3610 GCCAAATGATTTCCAGTTGGCAAGACCCAGAGCCAGCTGCCCCCT 3659
1193 GlyTrpThrGlnThrProThrProValProGlnsPlyGlyGlnProGln 1209
      |||
3660 CCATGGCGGGACAGACCAATGAGGTTTCAAGGATGAGAGAACCC.. 3707
1209 yGlnAspLeuArgHisProGlyThrSerLeuProAlaAspLeuProGlyA 1226
      |||
3708 .....AAGGGCCGGGAGAACCC..CACCTGCCCCCGA 3738
1226 rGpProProGlu.....ProCysHisProThrGlyThr 1236
      |||||
3739 GACCCAGCTCCAGGCTGCCCTTGTGCTCCCTGTGGCAGC 3779
```

100

seq\_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:AAS97178

seq\_documentation\_block:

ID AAS97178 standard; cDNA; 3561 BP.

XX AAS97178;

XX AC

XX AD

XX DT 26-FEB-2002 (first entry)

XX DE

Human metalloprotease partial DNA sequence #7.

KW Human; protease; PCR primer; cyostatic; immunomodulator; cardiacty;

KW vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;

KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;

KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;

KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;

KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;

KW immune-related disease; cardiovascular disease; neuronal disease;

KW migraine; sexual dysfunction; mood disorder; attention disorder;

KW cognition disorder; hypotension; hypertension; psychotic disorder;

KW dyskinesia; metabolic disorder; inflammatory disorder; ss.

XX OS

XX Homo sapiens.

XX PN

XX MO200183782-A2.

XX PD

XX 08-NOV-2001.

XX PF

XX 04-MAY-2001; 2001WO-US14431.

XX PR

XX 04-MAY-2000; 2000US-201879P.

XX PA

(SUGC-) SUGEN INC.

XX PI

XX Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;

XX PI

XX Payne V;

XX DR

XX MPI: 2002-041502/05.

XX P-PSDB; AAU72895.

XX PT

XX Novel protease polypeptide useful for screening for substances that may

XX PT

XX be used to treat, e.g., cancers, immune-related diseases,

XX PT

XX cardiovascular disease, migraine, pain, psychotic and inflammatory

XX PT

XX disorders -

XX PS

XX Claim 30: Figure 1K-L; 232pp; English.

XX CC

The invention relates to an isolated, enriched, or purified protease

XX CC

XX polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to

XX CC

XX screen for substances (S) that may modulate its activity. Administering

XX CC

XX S (which modulates protease activity in vitro) may be used to treat a

XX CC

XX disease or disorder selected from cancers (e.g., of tissues, of blood or

XX CC

XX haematopoietic origin, of the breast, colon, lung, prostate, cervical,

XX CC

XX brain, ovarian, bladder or kidney), immune-related diseases and

XX CC

XX disorders, cardiovascular disease, brain or neuronal-associated diseases

XX CC

XX (e.g., central or peripheral nervous system diseases, migraine, pain,

XX CC

XX sexual dysfunction, mood disorders, attention disorders, cognition

XX CC

XX disorders, hypotension, hypertension, psychotic disorders, neurological

XX CC

XX disorders and dyskinasias), metabolic disorders and inflammatory

XX CC

XX disorders. (I) may also be useful as a diagnostic tool for a disease or

XX CC

XX disorder such as those above. AAS97159-AAS97195 represent human

XX CC

XX protease coding sequences and primers of the invention.

XX SQ

Sequence 3561 BP; 694 A; 1131 G; 1113 G; 623 T; 0 other;

alignment\_scores:



```

|||||
1701 ATGGGGGGGGGGTGGCATCCCGCAGCCGGAGCTGCACACACCCCTCCC 1750
584 rolaIaYrGLyGLyArGProCysLeuGLyPrometPheGLuTrpGLInVal 600
1751 CAGCCATATGAGGCCCCCGTGTAGGCCCATGTTCAGTACCAAGTTC 1800
601 CysAsnSerGluGluCysProGlyThrTrpGLuAspPheAlaGlnGI 617
1801 TGCACACGAGGAGTGCCTCGGACCTACGAGACTTCGGGCCACACA 1850
617 nCySaLaLyAsrGAsnSerTrpValHisGlnAsnAlaLysHisSerT 634
1851 GTGTGCCAAGCCAACTCGTACTATGTGACCCAGATGCCAAGACAGCT 1900
634 rPValProGlyGluProAspAspAlaGlnLysCysGluLeuILecys 650
1901 GGGTGGCTAGAGCTGACGATGACGCCAGAACTGTGAGCTGATCTGC 1950
651 GlnSerAlaAspThrGLyAspValAlaPheMetAsnGlnValHisAs 667
1951 CAGTCGGCGGACACGGGGAGCTGTGTTCATGAACCAAGTGTTCACGA 2000
667 pGLyThrArGcysSerTrpArGAspProTrpSerValCysAlaArGLyG 684
2001 TGGGACAGCTGCACCTACCGGACCCATACAGCTGTGCGCGTGGCG 2050
684 lucYsValProValGLyCysAspLysGluValGLySerMetLysAlaAsp 700
2051 AGTGTGCTGTCTGCGCTGTGTACAAAGAGGTGGGTCTCATGAAGCGGAT 2100
701 AspLysCysGLyValCysGLyGLyAspAsnSerHisCysArGThrValLy 717
2101 GACAAAGTGTGAGTCTGCGGGGTGACAACTCCACCTGCAGAGACTGTGAA 2150
717 sGLyThrLeuGLyLysAlaSerLysGlnAlaGLyAlaLeuLysLeuValG 734
2151 GGGGACGCTGGGCAAGCCTCCAAAGCAGGAGGAGCTCTCAAGCTGTGTC 2200
734 lInLeProAlaGLyAlaArGHisILeGlnILeGluAlaLeuGluLysSer 750
2201 AGATCCCGCAGAGTGCAGGCACATCCAGATTGAGGCACCTGGAGAAATCC 2250
751 ProHisArGserValValLysAsnGlnValThrGLySerPheILeLeuAs 767
2251 CCCACCGCATTTGTGTGAAGAACAGCTCACCGCGACTTCATCTCTCAA 2300
767 nProLysGLyLysGluAlaThrSerArGThrPheThrAlaMetGLyLeuG 784
2301 CCCCAAGGCGCAAGGAGCCACCAAGCCGACCTTCACCGCATGGGCTGTG 2350
784 lUTrpGLuAspAlaValGLuAspAlaLysGluSerLeuLysThrSerGLy 800
2351 AGTGGAGGATGCGGTGAGAGATGCCAAGAAAGCCCAAGACACACGGG 2400
801 ProLeuProGluAlaILeAlaILeAlaLeuProProThrGluGly 817
2401 CCCCTGCTGAAGCCATTCCTCTGCTCTGCTCTCTCTCTCTCTCTCTCT 2450
817 yProArGSerSerLeuAlaTrpLysTrpValILeHisGluAspLeuLeuP 834
2451 CCCCCGAGAGCCTGGCTACCAAGTACGTCATCCATGAGGACCTGTCTGC 2500
834 roLeuILeGLySerAsnAsnValLeuLeuGluGluMetAspThrTrpGLu 850
2501 CCTTATCGGGAGCAACATATGCTCTCTGAGAGATGACACACATATGAG 2550
851 TrpAlaLeuLysSerTrpAlaProCysSerLysAlaCysGLyGLyGLy 867
2551 TGGGCGCTCAAGAGCTGGGCCCTCTGACGCAAGGCTGTGGAGAGGAGAT 2600
867 eGlnPheThrLysTrpGLyCysArGArGArGAspHisILeMetValG 884
|||||
2601 CCAGTTACACAAATAGCGCTGCCGGCCAGACGAGACCAACCATGTGTC 2650
884 lInArGHisLeuCysAspHisLysLysArGProLysProILeArGArGArG 900
2651 AGCGACACCTGTGTGACCAAGAAAGAGCCCAACCATCGCGCGGCGC 2700
901 CysAsnGlnHisProCysSerGlnProValTrpValThrGluGluTrpGL 917
2701 TGCACACACACCCGCTGTCTGAGCTGTGTGTGTGTGTGTGTGTGTGTG 2750
917 yAlaCysSerArGserCysGLyLysLeuGLyValGlnThrArGLyILeG 934
2751 TGGCTGCACCCGAGAGCTGTGGAAAGCTGGGGGTGCAGACAGCGGGATAC 2800
934 lncYsLeuLeuProLeuSerAsnGLyThrHisLysValMetProAlaLys 950
2801 AGTGCCTGATGCCCTCTCTCAATGGAACCCCAAGTATATCGCGCCAAA 2850
951 AlACysAlaGLyAspArGProGluAlaArGArGProCysLeuArGValPr 967
2851 GCCTGTGCGGGGACCGGCTGAGGCGCCGAGCGCCCTGTCTCCAGATGCC 2900
967 oCySProAlaGlnTrpArGLeuglyAlaTrpSerGlnCysSerAlaThrC 984
2901 CTGCCAAGCCAGTGGAGGCTGGAGCCTGGTCCAGTCTGTGCCACT 2950
984 ySGLyGLyGLyILeGlnGlnArGlnValValCysArGThrAspAlaAsn 1000
2951 GTGGAGAGGCGATCCAGCAGCGGCGAGTGTGTGCAGAGCAACACGCAAC 3000
1001 SerLeuGLyHisCysGLyGLyAspArGProAspThrValGlnValCysE 1017
3001 AGCTCTGGGCAATGTGGAGGGGATAGCCAGACACTGTCCAGAGCTGTGAC 3050
1017 rLeuProAlaCysGLyGLyAsnHisGlnAsnSerThrValArGAlaAspY 1034
3051 COTGCTGCTGTGTGA..... 3066
1034 aLTrpGLuLeuGLyThrProGluGLyGlnTrpValProGlnSerGLyPro 1050
3066 ..... 3066
1051 LeuHisProILeAsnLysILeSerSerMetCysAlaAlaGluProCysH 1067
3067 ..... 3080
1067 rGLyAspArGserValPheCysGlnMetGluValLeuAspArGTrpCys 1084
3081 GGGAGACAGGTCTGTCTGTGACAGATGGAAGTGCATCGCTACTGCT 3130
1084 erILeProGLyTrpHisArGLeucCysValSerCysILeLysLysAla 1100
1101 CCATTCGGGCTACACAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3180
1101 SerGlyProAsnProGlyProAspProGlyProThrSerLeuProProPh 1117
3181 TCGGGCCCAACCTGTGCCAGACCTGTGCCCAACTCTACTGCCCCCTT 3230
1117 eSerThrProGlySerProLeuProGlyProGlnAspProAlaAspAla 1134
3231 CTCACACTCTGGAAGCCCTTACCAAGAGCCCAAGAGCCTGAGATGTGCG 3280
1134 lAGluProProGLyLysProThrGLySerGluAspHisGlnHisGLyArG 1150
3281 CAAGAGCTCTGTGAAGGCCAAGCGGATCAGAGGACCATCAGCATGCGCGA 3330
1151 AlaThrGlnLeuProGlyAlaLeuAspThrSerSerProGlyThrGlnHis 1167
3331 GGCACACAGCTGCCAGAGCTCTGATACAACTCCCAAGGAGCAACCA 3380
1167 sProPheAlaProGluThrProILeProGlyAlaSerTrpSerILeSerP 1184
3381 TCCCTTTGGCCCTGAGACCAATCCCTGGAGCATCTGTGAGCATCTGCC 3430

```

1184 rothhrprogllygllyleuprotiprlytrpThrGlnThrProthrPro 1200  
 3431 CTACACACCCCGGGGGGCTGCTTGGGGCTGACACACACACACGCA 3480  
 1201 ValProGluAspLysGlyGlnProGlyGluAspLeuArgHisProGlyThr 1217  
 3481 GTCCCTGAGAGCAAGGCACTGAGAAAGACCTGACATCCCGGAC 3530  
 1217 rSerLeuProAlaAspLeuPro 1224  
 3531 CAGCTCCCTGCTGCTCCCTCCG 3552

seq\_name: /STDS/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH47048

seq\_documentation\_block:  
 ID AAH47048 standard; DNA; 3144 BP.

AC AAH47048;

29-OCT-2001 (first entry)

Larynx carcinoma associated protein-1 (LarCAP-1) encoding DNA.

Larynx carcinoma associated protein-1; LarCAP-1; carcinoma; cytostatic;  
 anti-metastatic; anti-arthritic; osteopathic; immunosuppressive; human;  
 immunostimulant; cerebroprotective; vasotropic; anti-inflammatory;  
 dermatological; cardiac; vaccine; ds.

OS Homo sapiens.

Location/Qualifiers

1..3144

CDs

/tag= a

/product= "LarCAP-1"

/note= "the stop codon is not indicated"

WO200159133-A1.

PD 16-AUG-2001.

PF 12-FEB-2001; 2001WO-EP01525.

PR 14-FEB-2000; 2000EP-0102955.

PA (MERE) MERCK PATENT GMBH.

PI Duecker K, Hentsch B, Hohnsels J, Frohne M;

DR WPI: 2001-483569/52.

P-PSDB: AAB85695.

Larynx carcinoma associated protein-1 polypeptide for the treatment of  
 carcinomas, metastasis, arthritis, osteoporosis, immune disorders,  
 stroke, ischemia, autoimmune disease, angiogenesis, skin disorders and  
 organ malfunctions -

Claim 5; Page 33-39; 47pp. English.

This DNA encodes a human Larynx carcinoma associated protein-1 (LarCAP-1)  
 polypeptide. The LarCAP-1 polypeptide can be expressed by standard  
 recombinant methodology. The LarCAP-1 polypeptide and polynucleotide are  
 useful in diagnostic assays and for the treatment of carcinomas,  
 metastasis, arthritis, osteoporosis, immune disorders, stroke, ischemia,  
 autoimmune disease, angiogenesis, skin disorders and organ malfunctions  
 especially heart hypertrophy.

Sequence 3144 BP; 633 A; 1000 C; 957 G; 554 T; 0 other;

alignment\_scores:  
 Quality: 5160.50 Length: 970  
 Ratio: 5.573 Gaps: 2

Percent Similarity: 95.464 Percent Identity: 95.361

alignment\_block:

US-09-938-330-20 x AAH47048 ..

Align seg 1/1 to: AAH47048 from: 1 to: 3144

291 ValAspGluIleuThrHisAspLysSerLeuGlyValHisIleAsnIleAl 307  
 238 GTAGATGAGATTATCCAGATGATGCTCTGGGGGTTCTATTAATATATG 287  
 307 AlaValArgLeuIleMetValGlyThrArgGlnSerLeuSerLeuIleG 324  
 288 CTTGCTCCGCTTATCATGTTGGCTACCGACAGTCCCTGAGACCTGATCG 337  
 324 LuArgGlyAsnProSerArgSerLeuGlnValCysArgTrpAlaHis 340  
 338 AGCGGGGAACCCCTCAGCAGCTGAGAGAGTGTGTGGGGGACAC 387  
 341 SerGlnGlnArgGlnAspProSerHisAlaGluHisHisAspHisValVa 357  
 388 TCCAGACAGCGCCAGAGACCCAGCCAGCTGAGACCATGACACGCTTGT 437  
 357 LpheLeuThrArgGlnAspPheGlyProSerGlyTrpAlaProValThrG 374  
 438 GTTCTCACCAGGAGAGATTGGCCCTCAGAGTATGACCCGTCAC 487  
 374 LysMetCysHisProLeuArgSerCysAlaLeuAsnHisGlnAspGlyPhe 390  
 488 GCATGTGTACCCCTGAGAGAGCTGTGCTCCCTACCATGAGATGGCTTC 537  
 391 SerSerAlaPheValIleAlaHisGluThrGlyHisValLeuGlyMetG 407  
 538 TCTTCACGCTTCGATGATGCTCATGAGACCGGCGCATGCTCGGCA 587  
 407 uHisAspGlyGlnGlyAsnGlyCysAlaAspGluTrpTrpSerLeuGlySerY 424  
 588 GCATGACGCTCAGGGGAATGCTGTGCAGATGAGACAGCCTGGCAGCG 637  
 424 AlMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTrpSerArg 440  
 638 TCAATGCGCCCTCTGTGTCAGAGCTGCTCCACCGCTCATGTCCTCCG 687  
 441 CysSerLysLeuGluLeuSerArgTrpLeuProSerTrpAspCysLeu 457  
 688 TGCAGCAAGCTGGAGCTCAGCCGCTACTCCCTCTACAGACGCTCCCT 737  
 457 uAspAspProPheAspProAlaTrpProGlnProProGluLeuProGlyI 474  
 738 CGATGACCCCTTTGATCTGCTGCTGGCCCAAGCCCAAGAGCTGCGGA 787  
 474 LeuAsnTrpSerMetAspGluGlnCysArgPheAspPheGlySerGlyTrp 490  
 788 TCAACTACTCAATGATGATGAGCAGTGCCTTGTGACTTGTGACAGTGG 837  
 491 GlnTrpCysLeuAlaPheArgTrpPheGluProCysLysGlnLeuTrpC 507  
 838 CAGACCTGCTTGGCACTTCAGACCTTTGAGCCCTGCAAGCAGCGTGGTG 887  
 507 sSerHisProAspAsnProTrpPheCysLysTrpLysGlyProProL 524  
 888 CAGCCATCCTTGCAACCCGCTACTTCTGCAAGACCAAGAGGGGCCCCG 937  
 524 euAspGlyThrGlnCysAlaProGlyLysTrpCysPheLysGlnHisCys 540  
 938 TCGATGGAGCTGAGTGTGACCCCGCAAGTGTGCTTCAAAAGTCACTGC 987  
 541 IleTrpLysSerProGluGlnThrTrpGlnAspGlyLysTrpSerSe 557  
 988 ATCTGGAAGTCCCGGAGCAGACATATGCGCAGATGAGGCTGAGACTC 1037  
 557 rTrpThrLysPheGlySerCysSerArgSerCysGlyGlyValArgS 574

```

1038 CTGAGCAAGTTGGTCATGTTCCGCTCATGTGGGGCGGGCTGCAT 1087
574 eTASeArSerCyAsnAsnProSerProAlaTyrGlyAlaPro 590
1088 CCGGACCGGAGCTGCAACACCCCTCCCGACCTATGGAGCCGCCG 1137
591 CysLeuGlyProMetPheGlyTyrGlnValCysAsnSerGlyuGlyPr 607
1138 TGCTTAGGGCCATGTTGAGTACCAAGTCTCGAACAGGAGAGTCC 1187
607 OGlyThrTyrGluAspPheArgAlaGlnGlyAlaLysArgAsnSer 624
1188 TGGACCTACGAGGACTTCGGGCGCCAGAGTGTGCCAAGCCCACT 1237
624 YrTyrValHisGlnAsnAlaLysHisSerTrpValProTyrGluPro 640
1238 ACTATGTGCACCAAGATGCCAACACAGCTGGTGGCTTACGAGCT 1287
641 AsPAspAlaGlnLysCysGluLeuLecysGlnSerAlaAspThrGly 657
1288 GATGAGCCCGAAGTGTGAGTGTGCTCCAGTCGGCGAGACCGGG 1337
657 PValValPheMetAsnGlnValValHisAspGlyThrArgCysSer 674
1338 CGTGTCTTCATACCAAGCTGTTCCAGTGGAGACGCTGCACCTAC 1387
674 rGAspProTyrSerValCysAlaArgGlyGluCysValProValGly 690
1388 GGGACCCATACAGCGTCTGCGCGTGGGAGTGTGCTGTCCGCTGT 1437
691 AsPlyGlnValGlySerMetLysAlaAspLysCysGlyValCysGly 707
1438 GACAAAGAGGTGGGTCCATGAAGCGCATGACAAAGTGTGAGTCT 1487
707 YGlyAspAsnSerHisCysArgThrValLysGlyThrLeuGlyLys 724
1488 GGGTGCACACTCCACTGCAGGACTGTAGAGGAGCGCTGGGCAAG 1537
724 eTlysGlnAlaGlyAlaLeuLysLeuValGlnIleProAlaGlyAla 740
1538 CCAAGGAGGAGGAGCTCTCAAGCTGTGCGAGATCCAGCAGTGCC 1587
741 HisIleGlnIleGlnAlaLeuGlnLysSerProHisArgSerVal 757
1588 CACATCCAGATTGAGGCACTGGAGAGTCCCGCACCGCATTTGGTGA 1637
757 sAsnGlnValThrGlySerPheIleLeuAsnProLysGlyLysGlu 774
1638 GAACCAAGTCAACCGGAGCTTCACTCAACCCCAAGGAGGAGAAC 1687
774 hTSeArArgThrPheThrAlaMetGlyLeuGluTyrGluAspAla 790
1688 CAAGCGGAGCTTCAACCGCATGGGCTGTGGAGTGGAGATGGCGTG 1737
791 AsPAlaLysGlnSerLeuLysThrSerGlyProLeuProGlnAlaIle 807
1738 GATGCCAAGGAAGCTTCAAGACCGGGGCCCTGCTGAGGCATATGC 1787
807 AlIleLeuAlaLeuProThrGlnGlyGlyProArgSerSerLeuAla 824
1788 CATCTGCTCTCCCGCCCACTGAGGCTGGGCCCGCAGCAGCTGGCT 1837
824 YrLysTyrValIleHisGlnAspLeuProLeuIleGlySerAsnAsn 840
1838 ACAAGTACGATCATCATGAGGACTGCTGCCCTTATCGGGAGCAACA 1887
841 ValLeuLeuGlnGluMetAspThrTyrGluTyrPAlaLeuLysSer 857
1888 GTGCTCCGAGAGATGAGACACTATGATGGCGCTCAAGAGCTGGGC 1937
857 aProCysSerLysAlaCysGlyGlyIleGlnPheThrLysTyrGly 874
1938 CCCTGAGCAAGGCTGTGGAGAGGATCCAGTTCACCAATATAGGCT 1987
874 ySArGArGArGArAspHisHisMetValGlnArgHisLeuCysAsp 890
1988 GCGGCGGACAGGAGACACCATATGTTGAGGAGACACTGTGTACCA 2037
891 LysLysArgProLysProIleArgArgArgCysAsnGlnHisProCys 907
2038 AAGAAGGGCCCAAGCCATCCCGCGCGCTGCAACCAACACCCCTG 2087
907 rGlnProValTyrPValThrGluGlnTyrProGlyLysSerArgSer 924
2088 TCAGCCCTGTGGGTGACGAGAGAGTGGGTGCTGACACCGGAGCT 2137
924 LysLeuGlnGlyValGlnThrArgGlyIleGlnCysLeuLeuPro 940
2138 GGAAGCTGGGGGTGCAGACACGGGGGATACAGTGCCTGCTGCC 2187
941 AsnGlyThrHisLysValMetProAlaLysAlaCysAlaGlyAsp 957
2188 AATGGAACCCAAAGGTCAATGCCGCAAGGCTCGCGGGGAGCCGC 2237
957 OGlnAlaArgArgProCysLeuArgValProCysProAlaGlnTyr 974
2238 TGAGGCCCGAGCGGCTGTCTCCAGTGTGCCCTGCCACCGCAGT 2287
974 eugLysAlaTyrPSerGln 979
2288 TGGAGCTGTGTCGCCAAGAACTTGTGAGCATGATGATGCCAGAC 2337
979 979
2338 CTGTGTACTAAGATGAGGAGAACCTAGCATCAACGACAGAGCTG 2387
980 980
2388 TGCCCTGTGCAGCCACAGTCTGTGCTGTGCACCTGTGGAGGG 2437
988 lGlnGlnArgGlnValValCysArgThrAsnAlaAsnSerLeuGly 1004
2438 TCCAGACGCGGAGGTGTGTGTCAGAGCAACGCAACAGCTTGGG 2487
1005 CysGlnGlyAspArgProAspThrValGlnValCysSerLeuPro 1021
2488 TCGAGGGGAGATAGGCGACAGACTGTCCAGTGTGCACCTGCCG 2537
1021 sGlyGlyAsnHisGlnAsnSerThrValArgAlaAspValTyrP 1038
2538 TGGAGGAATCACAGAACTCCAGGTGAGGGCCGATGTGGGAAC 2587
1038 lYrThrProGlnGlyGlnTyrPAlaProGlnSerGlyProLeuHis 1054
2588 GAGCGCCAGAGGGAGTGGGTGCCACAAATCGAACCCCTACATCC 2637
1055 AsnLysIleSerSerMetCysAlaAlaGluProCysThrGlyAsp 1071
2638 AACCAAGATATCATCA 2678
1071 rValPheCysGlnMetGluValLeuAspArgTyrCysSerIlePro 1088
2679 TGTCTTTCGCAGATGGAAGTGTGATCTCACTGCTCATTCACCG 2728
1088 YrHisArgLeuCysCysValSerCysIleLysLysAlaSerGlyPro 1104
2729 ACCACCGGCTGTGCTGTGTCTGCTGATCAAGAAAGCTTGGGCC 2778
1105 ProGlyProAspProGlyProThrSerLeuProProPheSerThr 1121
2779 CTTGGCCCAAGACCTGGCCCAACCTCACTGCCCTTCTCCACTG 2828
1121 YSerProLeuProGlyProGlnAspProAlaAspAlaGluProPro 1138
2829 AAGCCCTTACAGAGACCCAGAGACCTGACAGTGTGAGAGACCT 2878

```



```

1138  LysProHnGlySerGluAspHisGlnHisGlyArgAlaThrGlnLeu 1154
1139  |||||||
2879  GAAGGCCACGCGATCAGAGACCATCAGATGCGGAGCCACACAGCCTC 2928
1155  ProGlyAlaLeuAspThrSerSerProGlyThrGlnHisProHnAlaPr 1171
1156  |||||||
2929  CCAGGAGCTGTGATACAGATCCCGGAGGAGCCACCATCCCTTTGCCCC 2978
1171  cGluThrProHnProGlyAlaSerTrpSerTrpSerProHnThrProG 1188
1172  |||||||
2979  TGAGACACCAATCCCTGAGCATCTGAGCATCTCCCTACACACCCCG 3028
1188  LysGlyLeuProTrpGlyThrThrGlnThrProHnProHnAlaProGluAsp 1204
1189  |||||||
3029  GGGGCTGCTCTGGGCTGAGCTGAGACACCTAGCCGAGCTCCCTGAGGAC 3078
1205  LysGlyGlnProGlyGluAspLeuArgHisProGlyThrSerLeuProAl 1221
1206  |||||||
3079  AAAGGCGCAACCTGGAGAGACCTGAGCATCCGCGACACGCTCCCTGTC 3128
1221  AspLeuPro 1224
1222  |||
3129  TCCCTCCCGC 3138

```

seq\_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAH20225

seq\_documentation\_block:

ID AAH20225 standard; cDNA; 3618 BP.

XX AAH20225;

XX AC AAH20225;

XX D7 07-AUG-2001 (first entry)

XX DE Human ADAM-type metalloprotease MDR55-encoding cDNA, SEQ ID NO:11.

XX KW Human: MDR55; ADAM-type metalloprotease; drug screening;

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 1..3618

XX FT /tag= a

XX FT /product= "Human MDR55"

XX FT /note= "ADAM-type metalloprotease"

XX PN JP2001017183-A.

XX PD 23-JAN-2001.

XX PF 09-JUL-1999; 99JP-0196584.

XX PE 09-JUL-1999; 99JP-0196584.

XX PT (YAMA ) YAMANOUCHI PHARM CO LTD.

XX DR WPI; 2001-275950/29.

XX DR P-PSDB; AAB73550.

XX PT A new metal protease and its preparation for use as an anti-cancer and

XX PT anti-arthritis therapeutic -

XX PS Example 2; Page 19-20; 22pp; Japanese.

XX PS The invention relates to the novel human ADAM (A Disintegrin And

XX CC Metalloprotease)-type metalloproteases MDR54 (AAB73549) and MDR55

XX CC (AAB73550). The metalloproteases can be used for the treatment of

XX CC cancers and arthritis. The invention also relates to the genes encoding

XX CC MDR54 and MDR55, vectors and host cells containing the MDR54 or MDR55

XX CC genes, the recombinant production of MDR54 and MDR55, and antibody

XX CC specific for MDR54 or MDR55, and methods of screening for compounds

XX CC which modulate the activity of MDR54 and/or MDR55. The present

XX CC sequence represents cDNA encoding human MDR55.

```

XX
SQ Sequence 3618 BP, 1035 A; 802 C; 880 G; 901 T; 0 other;

alignment_scores:
    Quality: 3624.00      Length: 1198
    Ratio: 4.081          Gaps: 17
    Percent Similarity: 74.124      Percent Identity: 55.593

alignment_block:
US-09-938-330-20 x AAH20225 ..
Align seg 1/1 to: AAH20225 from: 1 to: 3618

```

```

37 AspTyrGlyValThrValProCysSerThrAspPheArgGlyArgPhe 53
38  ....
121 GAGTATGAGCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 170
122  ....
53 UserValValSerGlyProAlaAlaAlaSerAlaGlySerMetValV 70
171 TCCCATACCTCTTCT ..... 186
70 AlaThrProProThrLeuProArgHisSerHisLeuArg ..... 84
187  ....
85  .... ValAlaArgSerProLeuHisProGlyGlyThrLeuTrpPr 98
208 GCGAGGAGCAGCTGCTTCCCAACCT ..... 231
98 OGIYArgValGlyArgHisSerLeuTyrPheAsnValThrValPheGly 115
232  .... GAGCAGTGTGTTCTTTAACTACATCCAGCGCATTTGGA 265
115 YSGIleuHisLeuArgLeuArgProAsnArgArgValValProGly 131
266 AAGATTTCATCTGAGCTAAAGCCCAACCTGACATGATGATGATGATG 315
132 SerSerValGlyTrpGlnGlu ..... 138
316 GCTGTGTGTGAGTGGCATGAGCATCTGCTGCTGCTGCTGCTGCTGCTG 365
139  .... AspPheArgGluLeuPheA 145
366 TCCCATTAACAACCATCAACAGAGAGCTGCTGCTGCTGCTGCTGCTGCTG 415
145 TGIleuProLeuArgGlnGlyValTyrThrGlyValThrGlyMet 161
416 CAGAGCTTGGCAGCTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 465
162 ProGlyAlaAlaValAlaIleSerAsnGlyAspGlyLeuAlaGlyLeu 178
466 CCAGGAACTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 515
178 eArgThrAspSerThrAspPhePheIleGluProLeuGlnValGlyGln 195
516 AAAAGTGTATATGAAGTATTCATTTGAAACCTTGGAAAGAGTAAAC 565
195 InGluysGluAlaSerGlyArgThrHisValValTyrArgArgGluAla 211
566 AGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 615
212 ValGlnGlnGlnIleProAlaGluProAspGlyAspLeuHis ..... As 225
616 GTAGAACAGGCTCCATAGCATGTGCAAAAGCTTCCACTACAGAGAGTC 665
225 ngIuAlaPheGlyLeuGlyAspLeuProAspLeuGlnValGlyAla 242
666 GAGCTGGAAGGCTTGGATGATGATGATGATGATGATGATGATGATGAT 715
242 spGlnLeuGlyAspThrGluArgGlyArgGlnHisAlaGlySer 258
716 AGCAGTGAATGAACAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 765

```



```

259 TyrSerIleGluValLeuLeuValAspAspSerValValArgPheH1 275
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
766 TACAATATCAGAGTACTCGTGGAGATGGATGACTGTGGTCCGTTTCCA 815
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
275 sglYlySGluHISValGlnAsnTyrValLeuThrLeuMetAsnIleValA 292
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
816 TGGCAAGAGCAGCTCCAAACTACCTCCTGACCTAATGAACATTGTGA 865
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 spGluIleTyrHISAspGluSerLeuGluValHISIleAsnIleAlaLeu 308
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
866 ATGAAATTTACCATATGAGTCTCCGAGAGTCATATTAATGTGGTCTG 915
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 ValArgLeuIleMetValGlyTyrArgInsLeuSerLeuIleGluArg 325
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
916 GTGGCATGATATATCTGGAGATTCGCAAGTCCATACCTCATAGAAAG 965
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 gglYAsnProSerArgSerLeuGluGlnValCysArgTyrAlaHisSerg 342
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
966 GGGAAACCATCCAGAGCTTGGAGAAATGTGTGCTGGCGCTCCCAAC 1015
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 IngIleArgGlnAspProSerHISAlaGluHISHisAspHISValAlphe 358
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1016 ACCAAAGATCTGATCTCAACCACTGTGACACACATGACCATGCAATTTT 1065
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
358 IeuthrArgGlnAspPheGlyProSer.....GlyTyrAlaProva 372
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1066 TTAACCAAGCAAGACTTGGACCTGTGATGCAAGAGATATGCTCCAGT 1115
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 IThrGlyMetCysHISProLeuArgSerCysAlaLeuAsnHISGlnAspG 389
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1116 CACCGGCAATGTGTCTACGAGAGAGAGTTGACCTCAATCATAGAGATG 1165
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 IyPheSerSerAlaPheValIleAlaHISGluThrGlyHISValLeuGly 405
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1166 GTTTTTCATCTGCTTTTGTACTACCCATCAACGGGCCATGTGTGGGA 1215
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 MetGluHISAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeuG 422
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1216 ATGGAGCATGTATGAGCAAGCAACAGCTGTGTATGAGACTGCTAATGG 1265
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
422 ySerIleMetAlaProLeuValGlnAlaAlaPheHISArgPheHISTrps 439
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1266 AAGTGTCTAGTGGCTCTGTGGACAGAGCATTCATTCCTTGTGACTGT 1315
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
439 eraTgCysSerIySleuGluLeuSerArgTyrLeuProSerTyrAspCys 455
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1316 CCCGATGAGTGTCAAGAACTGAAAGATATATCCATTCTTGTGACTGT 1365
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
456 LeuLeuAspAspProPheAspProAlaTyrProGlnProGluLeuArg 472
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1366 CTCTGTATGACCCCTTTGATCATGATTTGGCTTAAACTCCAGAACTTCC 1415
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
472 oGlyTlIleAsnTyrSerMetAspGluGlnCysArgPheAspPheGlySerg 489
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1416 TGGATATCAATATATCTATGATGATGCAATGTCTTTTGTATTTGGTGG 1465
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
489 IyTyrGlnThrCysLeuAlaPheArgThrPheGluProCysLysGlnLeu 505
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1466 GCTATATAAAGTGCACCGCTTCCGAACCTTTGACCCATGTAACAGCTG 1515
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
506 TyrCysSerHISProAspAsnProTyrPheCysLysThrLysLysGlyPr 522
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1516 TGGGTATGACCTCTGATTAATCCCTACTTTTGTAAAGACATAAAAGGAGAC 1565
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
522 oProLeuAspGlyThrGluCysAlaProGlyLysTrpCysPheLysGlyH 539
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1566 TCCACTTGATGGAGCTGAATGTCTCTGTGAATAATGTCTATAAGGCTC 1615
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
539 IScySIIeTrrPlySerProGluGlnThrTyrGlyGlnAspGlyLysTyr 555
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1616 ATTGCAATGTGGAAGATGCTAATCAG...CAAAACAGATGCAATTTGG 1662
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

556 SerSerTrrPThrLysPheGlySerCysSerArgSerCysGlyGlyValA 572
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1663 GGGTCAATGAGCTAAATTTGGCTCTCTCTCGAGACATGTGGAAGCTGGGT 1712
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
572 IArgSerArgSerArgSerCysAsnAsnProSerProAlaTyrGlyValA 589
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1713 TCGTTTCAGAACACCGCAGTGCATATTAATCCATGCTCATGATGTGGTC 1762
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
589 rGProCysLeuGlyProMetPheGluTyrGlnValCysAsnSergGlu 605
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1763 AGCATTTGTCCTGTATTTTGTAGTACAGCTTTTGTAAACAGAAAGAA 1812
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
606 CysProGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaLysArgas 622
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1813 TGGCAAAAACACTTTGAGAGCTTCAGACACAGCAGTGTGACGAGGAAA 1862
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
622 nSerTyrTyrValHISGlnAsnAlaLysHISerTrrValProTyrClnP 639
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1863 CTCCCACTTTGATATCCAGAAATACCAACACCACTGCTGTGCTATATGAC 1912
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
639 roAspAspAlaGlnLysCysGluLeuLleCysGlnSerAlaAspThr 655
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1913 ATCTGACCCCAAGAAAGATGCCACCTTTTACTGTCAAGTCAAGAGACT 1962
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
656 GlYAspValValaPheMetAsnGlnValValHISAspGlyThrArgCysS 672
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1963 GGAGATGTTGCTTACATGATAAACCACTGTGCTATGATGAAGCAGCTGTT 2012
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
672 IYrArgAspProTyrSerValCysAlaArgGlyGlnCysValProValG 689
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2013 TTACAAATATCCATATATGACATATGTGTCCGAGAGAGTGTGTAAAGTGG 2062
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
689 IyCysAspLysGluValaLysSerMetLysAlaAspLysCysGlyVal 705
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2063 GCTGTGATTAAGAAATGCTTCTAATGAAGTTGAGCTAATGATGTGGTGC 2112
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
706 CysGlyGlyAspAsnSerHISCysArgThrValLysGlyThrLeuGly 722
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2113 TGTGGAGGAGATTAATTCCTGCTGCCAAGCTGAAGGGGACATTTACAG 2162
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
722 sAlaSerLysGlnAlaGlyAlaLeuLysLeuValGlnIleProAlaGlyA 739
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2163 AACGCCAGAGAACTTGGTACCTTAAGATGTTGATATATCCCTCGGGG 2212
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
739 IArgHISIleGlnIleGluAlaLeuGluLysSerProHISArgSerVal 755
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2213 CTAGACATGTGTATATCCAAAGAGAGGCTTCTCTCATATTTCTTGTCT 2262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
756 ValLysAsnGlnValThrGlySerPheIleLeuAsnProLysGlySgl 772
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2263 ATTAAGAACCAGGCTACAGGCAATTATATTTAAATGGCAAGGGAGAGA 2312
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
772 uAlaThrSerArgThrPheThrAlaMetGlyLeuGluTyrGluAspAla 789
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2313 AGCCAAAGTCCGCGACCTTATAGATCTTGGTGTGAGTGGATTAACA 2362
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
789 alGluAspAlaLysGlnSerLeuLysThrSergLysProLeuProGluAla 805
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2363 TTGAAGATGACATTGAAGCTTTCACACCAAGGACCTTTACATATATCT 2412
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
806 IleAlaIleLeuAlaLeuProProThrGluGlyGlyPyrArgSerSerg 822
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2413 GTTATTTGTTTGTATTA...CCTCAAGAAATGATACCCGCTTACGCT 2459
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
822 uAlaTyrLysTyrValIleHISGlnAspLeuLeuProLeuIleGlySera 839
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2460 GACATATTAATCATCATCATGACAGACTGTACTACATCAATCAACACA 2509
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
839 snAsnValLeuLeuGluGluMetAspThrTyrGluTrrAlaLeuLysSer 855
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2510 ACAATGTCAATCCAGGAAGATAGATACTTTTGAAGTGGCTTTGAAGAGC 2559
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
856 TrrAlaProCysSerLysAlaCysGlyGlyIleGlnPheThrLysTyr 872
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

|||||
2560 TGGTTCAGTGTTCACAAACCCCTGTGGTGGAGTTTCCAGTACCTTAATA 2609
872 TGLYSEGLUASPHISGLINISGLIYARGALATHRGILNLEUPROGLIYAL 889
2610 TGGATGCCGTAGAGAAAGTGATATAATAATGATCCATCGAGCTTGTGG 2659
889 SPHSISLVSATSPROLYSPROILEARGARGARGCYASANSGLINISPRO 905
2660 AGGCCACCAAAAACCCCAACCTATTAGACGATGTGCATATATCAAGAG 2709
906 CysSerGlnProValThrValThrGluGlnThrPGLYALAcysSerArgse 922
2710 TGTACATCATCTCTGTGTAGCAGAGAAATGGAAACACACGCAACAAAC 2759
922 TCGSGLYLSLGLYGLYALGlnThrArgGlyLGLNcysLeuLeuProL 939
2760 CTGTGGAGTGTCTGTATAGCTTCGACACTGTACGCTGCTGCTTCAGCCAC 2809
939 euSerAsnGlyThrHisLysValMetProAlaLysAlaCysAlaGlyAsp 955
2810 TCCTTGATGGACCAACCCGCTCTGTGCACAGCAAAATCTGCATGGGTGAC 2859
956 ArgProGluAlaArgArgProCysLeuArgValProCysProAlaGlnTr 972
2860 CGTCCCGAGACCGCCGCTGTACACAGATGCCCTGCGCTGCACACTG 2909
972 PArgLeuGlyAlaTrpSerGlnCysSerAlaThrCysGlyLGLNcylleg 989
2910 GAAGAAGAGACCCGTGAGTGAAGTGTTCAGTCACTCGGTGAGAGACGG 2959
989 LnglnArgGlnValAlaCysArgThrAsnAlaAsnSerLeuGlyHisCys 1005
2960 AGGTGAGGACAGCTCTCTGACAGGCTGGG.....GACCACTGT 2997
1006 GILGLYASPARPROASPTHRVALGlnValCysSerLeuProAlaCysG 1022
2998 GATGTTGAAGACCTGAGTGCCTGACAGCCCTGTCACTGCTCTTGA 3047
1022 YGLYASNHISGLINASerThrValArgAlaAspValITRGLIleuGlyT 1039
3048 TGA..... 3051
1039 hrProGluGlyGlnTrValProGlnSerGlyProLeuHisProIleAsn 1055
3051 ..... 3051
1056 LysIleSerSerMetCysAlaAlaGluProCysThrGlyAspArgSerVa 1072
3052 .....GACCAATGTTGGGAGACAAAGTCCAT 3077
1072 LPhecysGlnMetGluValLeuAspArgTyrCysSerIleProGlyTyrH 1089
8078 ATTCTGTCAAAATGAGAGTGTGGCAGCATACGCTCCATACCAGGTATA 3127
1089 LAsrLeuCysCysValSerCysIleLysLysAlaSerGly..... 1102
3128 ACAAGTATTGTTGTGAGTCTCCACCAAGCCGAGTACAGCAACCCCTGCACCA 3177
1103 .....ProAsnPr 1105
3178 CCATACCTTCTAGAGCTGTGAAACATCATGATGTCATCTCTAACC 3227
1105 oGlyProAspProGly.....ProThrSerLeuProPrope. 1117
3228 TAGTGAACCTCCCTAGATCTTAGTATGCTACATCTTGGTCTCTATAC 3277
1118 .....SerThrProGly.....SerProLeu 1124
3278 ATTTCAGAGACCCCTGCAAGAGAGTCTTTGAGTACGATCTCTCAAGTG 3327
1125 ProGlyProGlnAspProAlaAspAlaAlaGluProGlyLysProTh 1141
|||||

```

```

3328 GGAGTCCAAATGCATATGCT...GCTTTCAGGCCCAACAGTAAACCTGA 3374
1141 TGLYSEGLUASPHISGLINISGLIYARGALATHRGILNLEUPROGLIYAL 1158
3375 TGGTCTCAATTTACGCCAG.....AGGAGTGTCTAGCAGACAGAGAGATA 3418
1158 euAspThr.....SerSerProGlyLThrGln 1166
3419 AGACTGTGAGACTGTGTACACCGTACCATTCTCCCAACCCACAG 3462
seq_name: /SIND5/gcgdata/geneseq/geneseqn-emb1/NA1999.DMT:AAZ32056
seq_documentation block:
ID AAZ32056 standard; DMT: 5774 BP.
XX
AC AAZ32056;
XX
DT 10-JAN-2000 (first entry)
XX
DE Human METH2 related EST AB002364.
XX
KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW angiogenesis inhibitor; abnormal wound healing; inflammation;
KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW diabetic retinopathy; macula degeneration; haemangioma; detection;
KW arterial-venous malformation; immune deficiency; ss.
XX
OS Homo sapiens.
XX
PN W09937660-A1.
XX
PD 29-JUL-1999.
XX
PF 22-JAN-1999; 99WO-US01313.
XX
PR 23-JAN-1998; 98US-0072298.
PR 28-AUG-1998; 98US-0098539.
XX
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
XX
PI IrueLa-Arispe L, Hastings GA, Ruben SM;
XX
DR WPI: 1999-590684/50.
XX
PT New isolated metalloprotease thrombospondin polypeptides, useful for
PT treating hyperproliferative disorders, cancers or autoimmune disorders
PS
PS Disclosure: Page 423-427; 457pp; English.
XX
CC AAZ32000 and AAZ32001 encode, and AA49501 and AA49502 represent, human
CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
CC respectively. METH1 and METH2 have been found to be potent inhibitors of
CC angiogenesis both in vitro and in vivo. They can be used for treating
CC cancer and other disorders related to angiogenesis including abnormal
CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
CC endometrial bleeding disorders, diabetic retinopathy, some forms of
CC macula degeneration, haemangiomas, and arterial-venous malformations.
CC They may be useful in treating deficiencies or disorders of the immune
CC system, by activating or inhibiting the proliferation, differentiation,
CC or mobilisation (chemotaxis) of immune cells. The etiology of these
CC immune deficiencies or disorders may be genetic, somatic, such as
CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
CC toxins), or infectious. They can also be used to treat inflammatory
CC conditions, both chronic and acute conditions. The products can also be
CC used for detection and diagnosis. AAZ32002 to AAZ2080, and AA49503 to
CC AA49511 represent sequences given in the exemplification of the present
CC invention.
XX
SO Sequence 5774 BP; 1766 A; 1129 C; 1258 G; 1621 T; 0 other;

```

## alignment\_scores:

Quality: 3615.50 Length: 1274  
 Ratio: 3.956 Gaps: 22  
 Percent Similarity: 71.743 Percent Identity: 53.375

## alignment\_block:

US-09-938-330-20 x AA232056 ..

Align seg 1/1 to: AA232056 from: 1 to: 5774

```

37 AsptYrGlyValThrValProCysSerThrAspPheArgGlyArgPheLeu 53
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
110 GAGTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 159
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
53 uSerHisValAlaSerGlyProAlaAlaAlaSerAlaGlySerMetValV 70
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
160 CTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 175
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
70 aLAspThrProThrLeuProArgHisSerSerHisLeuArg..... 84
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
176 .....ValAlaArgSerProLeuHisProGlyGlyThrLeuTrpPr 196
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
197 GCGAGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 220
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
98 oGlyArgValGlyArgHisSerLeuThrPheAsnValThrValPheGlyL 115
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
221 .....GAGCAGCTGCTCTTAACTCACTCAGCGCATTTGGAA 254
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
115 ySgLeuLeuHisLeuArgLeuArgProAsnArgArgLeuValValProGly 131
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
255 AAGATTTCATCTGCGACTAAAGCCACACACTCAACTAGTACCTGCTGCG 304
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
132 SerSerValGlyTrpGlnGlu..... 138
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
305 GCTGTGTGGAGTGGCATGAGACATCTCTGCTGCTGGGAATATAACCGCA 354
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
139 .....AspPheArgGlyLeuLeuPhe 145
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
355 TCCCATTAACAACCATCAACAGAGAGCTACGATATGAAATCCGGAATA 404
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
145 rGlnProLeuArgGlnGlnCysValThrThrGlyValThrGlyMet 161
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
405 CAGAGCCTTTGACAGCTAACTGCTGCTATGTTGCTGACATCGTGACAT 454
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
162 ProGlyAlaAlaValAlaAlaSerAsnCysAspGlyLeuAlaGlyLeu 178
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
455 CCAGAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 504
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
178 eArgThrAspSerThrAspPhePheLeuProLeuGlnArgGlyGlnG 195
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
505 AAAAGTATTAATGAGAGATTTTCATTGAACCTTGGAAAGAGTAAAC 554
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
195 InGlyValGlnAlaSerGlyArgThrHisValValThrArgArgGlnAla 211
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
555 AGATGGAGGAGAAAGAAAGAGGAGTTCATGTTGCTGCTCAAGAGATCAG 604
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
212 ValGlnGlnGlnTrpAlaGluProAspGlyAspLeuHis..... 225
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
605 GTAGAACAGGCTCCCATGACATGATTCACAAAGACTTCCACTACAGACAG 654
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
225 nGluAlaPheGlyLeuGlyAspLeuProAsnLeuLeuGlyValGlyVal 242
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
655 GGACCTGGAAAGCCTTGATGATCTAGGTACTGTTTATGCAACATCACCAC 704
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
242 spGlnLeuGlyAspThrGlyArgGlyArgThrHisAlaLysProGlySer 258
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
705 AGCAGCTGAATGAAGAAATGAGAGCGCGCAGACACGCGGAGAGAAAGCAT 754
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
259 TyrSerIleGlyValLeuLeuValValAspAspSerValValArgPheH 275
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

755 TACAAATATCGAGTACTGCTGGAGTGGATGACTGCTGCTGCTGCTGCTCA 804
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
275 sGlyValGlnHisValGlnAsnThrValLeuThrLeuMetAsnIleValA 292
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
805 TGSCAAAGAGCAGCTGCTCAAAACTACTCTGCTGCTGCTGCTGCTGCTG 854
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
292 spGluIleGlyHisAspGlySerLeuGlyValHisIleAsnIleAlaLeu 308
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
855 ATGAATTTTACCATGATGAGTCCCTGCGAGTGCATATGAATGCTGCTG 904
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
309 ValArgLeuIleMetValGlyThrArgGlnSerLeuSerLeuIleGluAr 325
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
905 GTCCGATATTAATCTGGAGATATGCAAACTCCTCACTCCTCATAGAAAG 954
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
325 gGlyAsnProSerArgSerLeuGlnIleValCysArgTrpAlaHisSerG 342
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
955 GGGAAACCATCCAGAGCTTGAGAAATGTGTGTGCTGGGCGTCCCAAC 1004
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
342 InGlnArgGlnAspProSerHisAlaGlnHisAspHisValAlaPhe 358
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1005 ACCAAAGATCTGATCTCAACCTCTGAAACACATGACATGCAATTTT 1054
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
359 LeuThrArgGlnAspPheGlyProSer.....GlyThrAlaProVa 372
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1055 TTACACAGGCAAGACTTTGGACTGCTGCAATGCAAGAGATTTGCTCAG 1104
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
372 ThrGlyMetCysHisProLeuArgSerCysAlaLeuAsnHisGlnAspG 389
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1105 CACCGCATGTGTCATCCAGAGAGAGTTTACCCTGATATGATAGAGATG 1154
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
389 TyrPheSerSerAlaPheValIleAlaHisGlyThrGlyHisValLeuGly 405
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1155 GTTTTTCATCTGCTTTTGTAGTACGCCATGAAGGCGCATGCTGTTGGA 1204
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
406 MetGlnHisAspGlyGlnGlyAsnGlyCysAlaAspGlyThrSerLeuG 422
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1205 ATGAGACATGATGAGACAGAGCAGAGGTGTGTATGAGACTGTAAGGG 1254
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
422 ySerValMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTrpS 439
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1255 AAGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1304
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
439 eArgCysSerLeuGlnLeuSerArgThrLeuProSerTyrAspCys 455
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1305 CCGGATGACATGCTGCAAGAACTGAAGATATATCCATTCCTATGACTGT 1354
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
456 LeuLeuAspAspProPheAspProAlaTrpProGlnProProGlyLeuPr 472
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1355 CTCCCTGATGACCTTTTATGATCATGATTGGCGTAAACTCCCAAGACTT 1404
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
472 oGlyIleAsnThrSerMetAspGlnGlnCysArgPheAspPheGlySerG 489
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1405 TGGATATCATTAATTTCTATGATGAGCAATGCTGTTTATGATTTTGGT 1454
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
489 TyrGlnThrCysLeuAlaPheArgThrPheGlnProCysValGlnLeu 505
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1455 GCTATTAAGATGTCACCGCGTTCGAACTTTTACACCATTAACACGCTG 1504
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
506 TrpCysSerHisProAspAsnProTyrPheCysLysThrLysLysGlyPr 522
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1505 TGGTGTAGCCATCTCTGATTAATCCCTACTTTTGTAAAGACATAAAAGG 1554
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
522 oProLeuAspGlyThrGlnCysAlaProGlyLysTrpCysPheLysGlyH 539
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1555 TCCACTTGATGAGACTGAATGCTGCTGGAATATGCTCTTAAGAGGCTT 1604
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
539 iCysIleTrpLysSerProGlnGlnThrTyrGlyGlnAspGlyGlyTrp 555
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1605 ATTGCATGTGAAGAAATGCTAATCAG...CAAAACAAATATGCAATTCG 1651
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
556 SerSerTrpThrLysPheGlySerCysSerArgSerGlyGlyValGlyVa 572
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

1652 GGTGATGAGCTAAATTGGCTCTCTCGGACATGTGGAACGTGCT 1701  
 572 IArgSerArgSerArgSerCysAsnAsnProSerProAlaTyrGlyVal 589  
 1702 TCGTTTCAGAAACAGCCAGTGCATATATCCCATGCTCATCATGAGTGC 1751  
 589 rGpProCysLeuGlyProMetPheGluTyrGlnValCysAsnSerGluGlu 605  
 1752 AGGATTTGCTCCGTGTTAATTGATGACACGCTTTGTGTACACAGAA 1801  
 606 CysProGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaLysArgAs 622  
 1802 TGCCAAAACACTTGTGAGACTTCAGAGCACAGAGTGTGACAGCGAAA 1851  
 622 nSerTyrTyrValHisGlnAsnAlaLysSHSSTerTyrValProTyrGluP 639  
 1852 CTCCACCTTGAATACAGATACCAACACACTGCTGTGCATATATGAC 1901  
 639 roAspAspAlaGlnLysCysGluLeuLeuLeuLeuLeuLeuLeuLeu 655  
 1902 ATCCGACCCCAAGAAAGATGCCACTTACTGTCTGCTCAGTCCAGAGACT 1951  
 656 GlyAspValValPheMetAsnGlnValValHisAspGlyThrArgCysSe 672  
 1952 GAGATGTTGCTTACATGAACAACTGCTGATGATGGAACGCACTGCTC 2001  
 672 TTYrArgAspProTyrSerValCysAlaArgGlyGluCysValProValG 689  
 2002 TTACAAATATCATATAGCATATGTGTCGAGAGAGTGTGTAAGAGCTG 2051  
 689 IYcAspAspGluValGlySerMetLysAlaAspAspLysCysGlyVal 705  
 2052 GCTGTGATTAAGAAATTTGCTTCTAATTAAGTGTGAGATTAAGTGTGTC 2101  
 706 CysGlyGlyAspAsnSerHisCysArgThrValLysGlyThrLeuGlyLys 722  
 2102 TGTGAGAGGATTAATTTCCACTGCGCAACCGTGAAGGAGCATTTACAG 2151  
 722 sAlaSerLysGlnAlaGlyAlaLeuLysLeuValGlnLeuProAlaGlyVal 739  
 2152 AACCTCCAGAGAGCTTGGGTACCTTAAGATGTTGATATACCCCTGAGG 2201  
 739 IArgHisIleGlnIleGlnAlaLeuLeuLysSerProHisArgSerVal 755  
 2202 CTAGACATGTGTTATCCAGAAAGAGAGGCTTCTCATATCTCTGCT 2251  
 756 ValLysAsnGlnValThrGlySerPheIleLeuAsnProLysGlyLysGlu 772  
 2252 ATTAGAACACAGGCTACAGGCGCATATATTTTAAATGGCAAGGGAGGA 2301  
 772 uAlaThrSerArgThrPheThrAlaMetGlyLeuGlnTyrPgluAspAlaVal 789  
 2302 AGCCAAAGTGGGACCTTCATAGATCTTGTTGAGAGTGGGATTTAAC 2351  
 789 aGlnuAspAlaLysGlnSerLeuLysThrSerGlyProLeuProGluAla 805  
 2352 TTGAAGATGACATTTGAAGCTTTACACCGAGTACGACTTTACATGATCCT 2401  
 806 IleAlaIleLeuAlaLeuProProThrGluGlyGlyProArgSerSerIle 822  
 2402 GTTATGTTTGAATTA...CTCAAGAAATGATACCCGCTCTAGCCT 2448  
 822 uAlaTyrLysTyrValIleHisGlnAspLeuLeuProLeuIleGlySer 839  
 2449 GACATTTAAGTACATCATGACATGAAGACTCTGACCTCAATCAACAGCA 2498  
 839 snAsnValLeuLeuGluGluMetAspThrTyrGluTyrAlaLeuLysSer 855  
 2499 ACAATGTCACTCAGAAAGATTAAGATCTTTGAGTGGGCTTTGAAGAGC 2548  
 856 ThrAlaProCysSerLysAlaCysGlyGlyGlyIleGlnPheThrLysTyr 872  
 2549 TGTCTCAGAGTTTCCAAACCTGTGTGTGAGGTTTCCAGTACACTTAATA 2598

872 rGlyCysArgArgArgAspHisIleMetValGlnArgHisLeuCys 889  
 2599 TGGATGCCCTAGAAAAGTAAATTAATGTTCATGCAAGCTCTGTG 2648  
 889 sPHisLysLysArgProLysProIleArgArgArgCysAsnGlnHisPro 905  
 2649 AGGCCAACAAAAGCCGAAACCTATTAGACAAATGTGCAAAATTCMAAG 2698  
 906 CysSerGlnProValTyrValThrGluGluGluGluGluGluGluGlu 922  
 2699 TGTACACATCTGCTGGGTAGCAGAGATGGAACTGCACCAAAAC 2748  
 922 rCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysLeuLeuProL 939  
 2749 CTCTGGAATCTGCTGATACACTTGCACCTGACCTGACGCTTACAGCAG 2798  
 939 euSerAsnGlyThrHisLysValMetProAlaLysAlaCysAlaLysP 955  
 2799 TCTTGATGCGACCAACCCCTGCTGTGCACAGCAAAATACATGAGTGA 2848  
 956 ArgProGluAlaArgArgProCysLeuArgValProCysProAlaGlnTr 972  
 2849 CGTCCGAGAGCGCGCGCTGTAAACAGATGCCCTGCCCTGCACAGT 2898  
 972 parGluGlyAlaTyrSerGlnCysSerAlaThrCysGlyGluGlyIleG 989  
 2899 GAAACAGAGCCCTGAGTGTGATGATGATGATGATGATGATGATGATG 2948  
 989 ILeuArgGlnValValCysArgThrAsnAlaAsnSerLeuLysHisCys 1005  
 2949 AGCTGAGGCGAGTCTCTGCGAGGCTGGG...GACCCTGT 2986  
 1006 GlnGlyAspArgProAspThrValGlnValCysSerLeuProAlaCysGlu 1022  
 2987 GATGCGAAAGCCTGAGTGTGATGATGATGATGATGATGATGATGATG 3036  
 1022 yGlnAsnHisGlnAsnSerThrValArgAlaAspValTyrPgluLeuGlyT 1039  
 3037 TGAT... 3040  
 1039 hrProGluGluGlnTyrValProGlnSerGlyProLeuHisProIleAsn 1055  
 3040 ..... 3040  
 1056 LysIleSerSerMetCysAlaAlaGluProCysThrGlyAspArgSerVal 1072  
 3041 .....GAAACATGTTGGAGACAAAGTCCAT 3066  
 1072 ILeuCysGlnMetGluValLeuAspArgGlyCysSerIleProGlyTyrH 1089  
 3067 ATTCTGCAATGAGAGTGTGGCAGCATCTCTCCATACAGGTTATA 3116  
 1089 IAspGluCysCysValSerCysIleLysLysAlaSerGly..... 1102  
 3117 ACAAGTTATGTGTGAGTCTCTGACAGCAAGGAGTACACCCGCAACCA 3166  
 1103 .....ProAsnPr 1105  
 3167 CCATACCTCTAGAAAGCTGTAAGAACTCATGATGATGATGATGATGATG 3216  
 1105 OGlyProAspProGly.....ProThrSerLeuProProPhe 1117  
 3217 TAGGACCTCCCTGATCTGATGATGATGATGATGATGATGATGATGATG 3266  
 1118 .....SerThrProGly.....SerProLeu 1124  
 3267 ATTAGAGACCCCTGCAAGAGATGCTTTGAGTACATCTCTTCAGTG 3316  
 1125 ProGlyProGlnAspProAlaAspAlaIleGluProProGlyLysProThr 1141  
 3317 GGAGTCCAAATGATATGCT...GCTTTCAGGCGCAAAACAGTAAACCTGA 3363

```

1141 rglYserGluAspHisGlnHisGlyArgAlaThrGlnLeuProGlyAlaL 1158
|||||:|||||:|||||:|||||:
3364 TGGTGGTAAATTACGCCAG.....AGAGCTGCTCAGCAAGCAGAAAGTA 3407
|||||:|||||:|||||:|||||:
1158 euAspThr.....SerSerProGlyThrGlnHisPro 1168
|||||:|||||:|||||:|||||:
3408 AGAGCTGTAGAGCTGTACCCGTACCATCTCCGCCACCACCAAGAGGCTC 3457
|||||:|||||:|||||:|||||:
1169 .....PheAlaProGly 1172
|||||:|||||:|||||:|||||:
3458 CACCTAGTTCAGCTTCACAAATGCTGCTCTCTTTGACAGCCAG 3507
|||||:|||||:|||||:|||||:
1172 uThrProIleProGlyAlaSerThrSerIleSerProThrThrProGly 1189
|||||:|||||:|||||:|||||:
3508 TGATTCATAGAGTGTCTTCTTCAGCAAGAACTCAAGAAAGATGAA 3557
|||||:|||||:|||||:|||||:
1189 lYLeu.....ProTrpGlyTrpThrGlnThrProThr... 1199
|||||:|||||:|||||:|||||:
3558 AGATCATTGTACAAACAGACCTCG.....ACAAATCATCCACCTTA 3598
|||||:|||||:|||||:|||||:
1200 .....ProValPro.....GluAspLysGlyGlnProGly 1209
|||||:|||||:|||||:|||||:
3599 GAAAGATGAGAAAGTGAACCAAAAGGCTAGAAACAGAGAAACCTGG 3648
|||||:|||||:|||||:|||||:
1209 yGluAspLeuArgHisProGly 1216
|||||:|||||:|||||:|||||:
3649 ACAACCTCTCTCTTCCATGGT 3670
|||||:|||||:|||||:|||||:

```

seq\_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAA95824

seq\_documentation\_block:

ID AAA95824 standard; cDNA; 5774 BP.

AC AAA95824;

DT 23-FEB-2001 (first entry)

DE Human metalloproteinase KIAA0366 cDNA.

KW Human: KIAA0366; ADAMTS; metalloproteinase; ADAM;

KW a disintegrin and metalloproteinase domain; thrombospondin domain;

KW vacuole; notropic; neuroprotective; antiparkinsonian;

KW cerebroprotective; cytoskeletal; antihypertic; immunosuppressive;

KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;

KW autoimmune disease; brain tumour; brain injury; ss.

XX Homo sapiens.

Key Location/Qualifiers

CDS 2..3607

/\*tag= a

/partial=

/product= "KIAA0366"

WO200053774-A2.

14-SEP-2000.

08-MAR-2000: 2000WO-US06237.

08-MAR-1999: 99US-0264585.

(NEUR-) NEUROCRINE BIOSCIENCES INC.

PI Kerner GS, Clark M, Maki R;

DR MPI: 2000-594326/56.

DR P-PSDB: AAB21254.

XX Polynucleotide encoding novel members of a disintegrin,

PT metalloproteinase and thrombospondin domain protein family used to

PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -

XX

PS Example 3; Fig 7; 129pp; English.

XX The present sequence encodes human metalloproteinase KIAA0366. KIAA0366  
CC shows homology to the ADAMTS family of proteins. ADAMTS  
CC thrombospondin, integrin and metalloproteinase domains. ADAMTS  
CC polypeptides are useful for the manufacture of medicaments for treating  
CC conditions associated with neuroinflammation and/or neurodegeneration,  
CC such as Alzheimer's disease, Parkinson's disease and stroke. They are  
CC also useful for treating conditions associated with cell proliferation,  
CC cell migration, inflammation and/or angiogenesis, such as cancer,  
CC arthritis and autoimmune diseases. They can be used to treat patients  
CC afflicted with an invasive tumour, a brain tumour or brain injury.

Sequence 5774 BP; 1766 A; 1129 C; 1258 G; 1621 T; 0 other;

alignment\_scores:

Quality: 3615.50 Length: 1274

Ratio: 3.956 Gaps: 22

Percent Similarity: 71.743 Percent Identity: 53.375

alignment\_block:

US-09-938-330-20 x AAA95824 ..

Align seg 1/1 to: AAA95824 from: 1 to: 5774

```

37 AspTYrGlyValThrValProCysSerThrAspPheArgLysPheLe 53
|||||:|||||:|||||:|||||:
110 GAGTATGAGCTGTGTGCTCAGTCAAGCAAAATCTAGAGAGAGCTATCT 159
|||||:|||||:|||||:|||||:
53 userHisValValSerGlyProAlaAlaAlaSerAlaGlySerMetVal 70
|||||:|||||:|||||:|||||:
160 CTCCCATACTCTTCT..... 175
|||||:|||||:|||||:|||||:
70 aAspThrProProThrIleuProArgHisSerSerHisLeuArg..... 84
|||||:|||||:|||||:|||||:
176 .....ValAlaArgSerProLeuHisProGlyLysThrLeuTrpPr 98
|||||:|||||:|||||:|||||:
85 .....ValAlaArgSerProLeuHisProGlyLysThrLeuTrpPr 98
|||||:|||||:|||||:|||||:
137 GCGAGGCGCTGTCTCTCCAACT..... 220
|||||:|||||:|||||:|||||:
98 OGlyArgValGlyArgHisSerLeuTyPheAsnValThrValPheGly 115
|||||:|||||:|||||:|||||:
221 .....GAGCAGTGTCTTTAATCATCAGCGCATTTGGAA 254
|||||:|||||:|||||:|||||:
115 yGluLeuHisLeuArgLeuArgProAsnArgLeuValProGly 131
|||||:|||||:|||||:|||||:
255 AAGATTTCATCTGCGACTAAAGCCCAACACCTCACTAGTACCTGGG 304
|||||:|||||:|||||:|||||:
132 SerSerValGluTrpGlnGlu..... 138
|||||:|||||:|||||:|||||:
305 GCTGTGTGGAGTGGCATGAGACATCTCTGCTCCGGAATTAACGA 354
|||||:|||||:|||||:|||||:
139 .....AspPheArgLysLeuPhe 145
|||||:|||||:|||||:|||||:
355 TCCCATTAACACCATCAACAGAGAGTCTTACGTATAGAAATCCGAA 404
|||||:|||||:|||||:|||||:
145 rglInProLeuArgGlnGlyValTyThrGlyValThrGlyMet 161
|||||:|||||:|||||:|||||:
405 CAGAGCTTGTGCAGACTAAGTGTCTTGTGTGATGATGAGACAT 454
|||||:|||||:|||||:|||||:
162 ProGlyAlaAlaValAlaIleSerAsnGlyLeuAlaGlyLeu 178
|||||:|||||:|||||:|||||:
455 CCAGAGACCTCTGTGGCATCAGCAACTGTGATGCTGTGGATGAT 504
|||||:|||||:|||||:|||||:
178 eArgThrAspSerThrAspPheIleGluProLeuGluArgGlyGln 195
|||||:|||||:|||||:|||||:
505 AAAAAGTATATAGAGAGATTTTCACTGACACCTTGAAGAAAGTAA 554
|||||:|||||:|||||:|||||:
195 lngLysGluAlaSerGlyArgThrHisValValTyArgArgLys 211
|||||:|||||:|||||:|||||:
555 AGATGAGGAGAAAGAAAGAGATTCATGTGTCTACAAAGATCAGCT 604
|||||:|||||:|||||:|||||:

```



```

2402 GTTATGTTGATTATA...CCTCAGAAATATGTAACCCCTCAGGCT 2448
822 ualaTyrIleValIleHisGluAspLeuProLeuIleGlySerA 839
2449 GACATATAGTACATCCATCAGAAAGACTGTGACCTACAAATCAACAGCA 2498
839 sPaAsValLeuLeuGluMetAspThrIleGluTrpAlaLeuLeuSer 855
2499 ACAATGTCTCCAGAGAGATAGTACTTTGAGTGGGCTTTCAGAGGC 2548
856 TrpAlaProCysSerLysAlaCysGlyIleGlnPheThrIleTyr 872
2549 TGGTCTCAGGTTCCAAACCTGTGTGAGGTTCCAGTACCTAAATA 2598
872 rGlyCysArgArgArgAspHisMetValGlnArgHisLeuGlyS 889
2599 TGGATGCCCTAGAAAGTATATAAATGGTCCATGCGAGCTTCGTG 2648
889 sPHisLysLysArgProLysProIleArgArgArgCysAsnGlnHis 905
2649 AGGCCAACAAAAAGCCGAACCTATTAGACGATGTGCAATATTCAGAG 2698
906 CysSerGlnProValTrpValThrGluGlnTrpGlyAlaCysSerA 922
2699 TGTACACATCCACTGTGGTAGCAAGAAATGGTCCAGTCCACCAAA 2748
922 rCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysLeuPro 939
2749 CTGTGAGAGTTCTGCTATCAGCTTCGACAGTACGCTGCTTCAGC 2798
939 euSerAsnGlyThrHisLysValMetProAlaLysAlaCysAlaG 955
2799 TCCATTGAGGCAACACGCTCTGTCACAGCAAAATGCTGATGGG 2848
956 ArgProGluAlaArgArgProCysLeuArgValProCysProAlaG 972
2849 CGTCCCGAGAGCCCGCGCCCTGTAACAGAGTGCCTGCTCCACAG 2898
972 parGluGlyAlaTrpSerGlnCysSerAlaThrCysGlyIleGly 989
2899 GAAAGACAGACCTGAGTAGTGTCTCAGTGCCTGGGTGAAGAACG 2948
989 GlnGlnArgGlnValCysArgThrAsnAlaAsnSerLeuGlnHis 1005
2949 AGGTGAGGCAAGTCTCTGCAAGGCTGG...GACCACTGT 2986
1006 GluGlyAspArgProAspThrValGlnValCysSerLeuProAla 1022
2987 GATGTGAAAAAGCCTGAGTGGGTGACAGCTTCAACTGCTCTGTAA 3036
1022 yGlyAsnHisGlnAsnSerThrValArgAlaAspValTrpGluLe 1039
3037 TGAT..... 3040
1039 hrProGluGlyGlnTrpValProGlnSerGlyProLeuHisProI 1055
3040 ..... 3040
1056 LysIleSerSerMetCysAlaAlaGluProCysThrGlyAspArg 1072
3041 .....GAAACCATGTTTGGGAGCAAGGCCAT 3066
1072 lPheCysGlnMetGluValLeuAspArgTyrCysSerIleProGly 1089
3067 ATTCTGCAATGGAAGTGTGGCAGCATGCTCCATACAGGTTATA 3116
1089 lAsArgLeuCysCysValSerCysIleLysLysAlaSerGly..... 1102
3117 ACAAGTTATTTGTGAGTCTCTGACGACAGCAGTACCCCTGCACCA 3166
1103 .....ProAsnPr 1105

```

```

3167 CCATACCTTCTAGAAAGCTGTGAACATCATGATGATCATCTTAACCC 3216
1105 ocGlyProAspProGly.....ProThrSerLeuProProHe. 1117
3217 TGTGACCTCCCTAGATATCTCTAGTATGCTTACATCTTTGGTCTTA 3266
1118 .....SerThrProGly.....SerProLeu 1124
3267 ATTGAGAGACCCCTCAAGAAAGATGTCTTGTAGTACATCTTCTAG 3316
1125 ProGlyProGlnAspProAlaAspAlaAlaGluProProGlyLys 1141
3317 GGAGGTCCAAATGCATATGCT...GCTTTCAGGCCAAACGTAAACCTGA 3363
1141 rGlySerGlnAspHisGlnHisGlyArgAlaThrGlnLeuProGly 1158
3364 TGGTCTAATATTAGCCAG.....AGGAGTCTCAGCAGCAGCAGGA 3407
1158 euAspThr.....SerSerProGlyThrGlnHisPro 1168
3408 AGACTGTGAGACTGTGTCACGCTACCATCTCTCCACCCACAGAGG 3457
1169 .....PheAlaProG 1172
3458 CACCTCAGTTCAGCTTCACAAATGGCTGCTGCTCTTCTTTCAGCAG 3507
1172 uThrProIleProGlyAlaSerTrpSerIleSerProThrThrPro 1189
3508 TGATTCAATAGTGTCTTCTCTCAGCAGAACCTCAAGAAAGATGAA 3557
1189 lYleu.....ProThrPlyThrPheGlnThrProThr... 1199
3558 AGATCATTTGACACAGACGCTCG.....ACAAATATATCCACTTA 3598
1200 .....ProValPro.....GluAspLysGlyGlnProG 1209
3599 GAAAGATGAGAAAGTGAACCAAAAGCTAGAAACAGAGAGAAACCT 3648
1209 yGluAspLeuArgHisProGly 1216
3649 ACAACCTCTCTCTTCCATGCT 3670
seq_name: /SID5/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAAC90313
seq_documentation_block:
ID AAC90313 standard; DNA; 5774 BP.
AC AAC90313;
XX
DT 19-MAR-2001 (first entry)
XX
DE AB002364 cDNA clone.
XX
KW METH; metalloproteinase; thrombospondin; angiogenesis inhibition;
cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;
KW plaque neovascularisation; telangiectasia; haemophilic joint; EST;
KW angiofibroma; fibromuscular dysplasia; expressed sequence tag;
KW Crohn's disease; atherosclerosis; birth control; ss.
XX
OS unidentified.
XX
PN WO200071577-A1.
XX
PD 30-NOV-2000.
XX
PF 25-MAY-2000; 2000WO-US14462.
XX
PR 25-MAY-1999; 99US-0318208.

```







1105 CACGGCATGTGTATCCAGTAGAAGTTGTACCTGCAATCATGAGATG 1154  
389 LpHeserSerAlaPheValIleAlaHisGluThrGlyHisValLeuGly 405  
1155 GTTTTTCATCTGCTTTGTAGTAGCCATGAAGCGGCATGTGTGGGA 1204  
406 MetGluHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeuGly 422  
1205 ATGAGCATGATGAGAACAGGCAAGGTGTGTGTATGAGACTGCTATGGG 1254  
422 ySerValMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTrpS 439  
1255 AAGTGCATAGCTCCCTTGTGTCAAGCAGCATTCATCTTACACATGTGT 1304  
439 eArgCysSerLysLeuGluLeuSerArgTyrLeuProSerTyrAspCys 455  
1305 CCCGATGCAGTGTCAAGAACTGAAAGATATATCCATCTCTAAGACTGT 1354  
456 LeuLeuAspAspProPheAspProAlaTrpProGlnProGluLeuPr 472  
1355 CTCCTTGATGACCTTTTGATCATGATGTGGCTAAACCTCCAGAACTTC 1404  
472 oGlyLysAsnTyrSerMetAspGluGlnCysArgPheAspPheGlySerG 489  
1405 TGGATTCATATTATCTATGATGAGCAATGTCTTTGATTTGTGTGTG 1454  
489 lYTrGlnThrCysLeuAlaPheArgThrPheGluProCysLysGlnLeu 505  
1455 GCTATAAATGTGCAACCGCTTCCGAACTTTGACCATGATAAACACACTG 1504  
506 TrpCysSerHisProAspAsnProTyrPheCysLysThrLysLysGlyPr 522  
1505 TGGTGTACCATCTGATTAATCCCTACTTTGTAAAGACTAAAGAGAGCC 1554  
522 oProLeuAspGlyThrGluCysAlaProGlyLysTyrCysPheLysGlyH 539  
1555 TCCACTTATGAGACTGATGATGTCTGCTGGAATAATGTGTATTAAGGCTC 1604  
539 lScyLierTrpLysSerProGluGlnThrTyrGlyGlnAspGlyGlyTrp 555  
1605 ATTCGATGTGGAAGATGCTAATCAG...CAAAACAAGATGGCAATGTG 1651  
556 SerSerTrpThrLysPheGlySerCysSerArgSerCysGlyGlyIya 572  
1652 GGGTATGACTAAATTTGGCTCTGTTCTCGGACATGTGAACGTGTGTC 1701  
572 lArgSerArgSerArgSerCysAsnAsnProSerProAlaTyrGlyGlyA 589  
1702 TCGTTTCAGAACACAGCCAGTGCATTAATCCCATGCCCATCAATGTGTC 1751  
589 rArgProCysLeuGlyProMetPheGluTyrGlnValCysAsnSerGluGlu 605  
1752 AGGATGTCTCTGGTGAATTTTGTAGTACAGCTTTGTAAACAGAGAA 1801  
606 CysTrpGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaLysArgAs 622  
1802 TGCCAAAAAACACTTTGAGACTTCAGAGCACACAGCTGTGACAGAGAAA 1851  
622 nSerTyrTyrValHisGlnAsnAlaLysHisSerTrpValProTyrGluPr 639  
1852 CTCACACTTTCGAAATACAGAAATACCAACACCTGTTGCCATATGAAC 1901  
639 roAspAspAlaGlnLysCysGluLeuLeuLeuCysGlnSerAlaAspThr 655  
1902 ATCCGTGACCCCAAGAAAGATGACCTTTACTGTACGTCCAAAGAGACT 1951  
656 GlYAspValValPheMetAsnGlnValHisAspGlyThrArgCysSe 672  
1952 GGAAGATGTTCTTACATGAACAACGTGTGCTATGATGAGAACGCACTGTC 2001  
672 rTyrArgAspProTyrSerValCysAlaArgGlyLysCysValProValG 689  
2002 TTACAAAGATCCATATAGCATATGTGTGCGAGAGAGTGTGTGAAGTGG 2051

689 lYcysAspLysGluValAlGlySerMetLysAlaAspAspLysCysGlyVal 705  
2052 GCTGTATGAAGAATGTGTTCTATATAGTTGAGGATGAATGAATGTGTC 2101  
706 CysGlyGlyAspAsnSerHisCysArgThrValLysGlyThrLeuGlyLys 722  
2102 TGTGAGGAGATTAATCCACACTGCCGAACCGTGAAGGGGACATTTACAG 2151  
722 sAlaSerLysGlnAlaGlyAlaLeuLysLeuValGlnLeuProAlaGlyA 739  
2152 AACTCCAGAGAGCTGGGTACTTAAGATGTTGATATACCCCTGGGG 2201  
739 lArgHisIleGlnIleGluAlaLeuGluLysSerProHisArgSerVal 755  
2202 CTAGACATGTGTTAATCCAGAGAGAGGCTCTCTCCATATCTTGTCT 2251  
756 ValLysAsnGlnValThrGlySerPheIleLeuAsnProLysGlyLysGly 772  
2252 ATTAAGAACACAGGCTACAGGCCATTAATTTAAATGGCAAGGGGAGGA 2301  
772 uAlaThrSerArgThrPheThrAlaMetGlyLeuGlnIleTrpGluAspAlaV 789  
2302 AGCCAAAGTCGGGAGACTTACATAGATCTTGTGTGAGATGGATATATACA 2351  
789 aGluAspAlaLysGluSerLeuLysThrSerGlyProLeuProGluAla 805  
2352 TTGAAGATGACATTTGAAGTCTTCACACCGATGAGACTTATCATGATCT 2401  
806 lIleAlaIleLeuAlaLeuProProThrGlnGlyLysProArgSerIle 822  
2402 GTTATGTTTGTGATTA...CTCAAGAAAATGAATACCGCTGTAGCT 2448  
822 uAlaTyrLysTyrValIleHisGluAspLeuLeuProLeuIleGlySera 839  
2449 GACATATGAATACATCATCATGAGACCTGTGATACCTACATCAACAGCA 2498  
839 sAsnValLeuLeuGluGluMetAspThrTyrGluThrAlaLeuLysSer 855  
2499 ACAATGTCATCCAGGAAGATGATTAATCTTTGATGAGGGGCTTTGAAGAG 2548  
856 TrpAlaProCysSerLysAlaCysGlyGlyIleGlnPheThrLysTyr 872  
2549 TGTGTCAGGTTTCCAAACCTGTGTGGAGGTTTCCAGTACACTAAATA 2598  
872 rGlyCysArgArgArgArgAspHisHisMetValGlnArgHisLeuCysA 889  
2599 TGCATGCCGTAGGAAGATGATTAATAAATGTGCATTCGACGCTGTG 2648  
889 sPHisLysLysArgProLysProIleArgArgArgCysAsnGlnHisPro 905  
2649 AGGCCAACAAAGAACCGCAACCTATTAGAGAAATGTCAATTTCAAGAG 2698  
906 CysSerGlnProValTrpValThrGluGluTrpGlyAlaCysSerArgSe 922  
2699 TGTACATCATCCTCTGTGTAGCAGAAAGATGGAAACACTGCAACCAAA 2748  
922 rCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysLeuLeuProL 939  
2749 CTGTGAGAGTTCTGCTATCAGCTTCGACGTGACCTCTCTTACAGCAC 2798  
939 eSerAsnGlyThrHisLysValMetProAlaLysAlaCysAlaLysP 955  
2799 TCTTGTATGACCAACACCGCTGTGTACAGCAAGAAATCTGCTGGGTAGC 2848  
956 ArgProGluAlaArgArgProCysLeuArgValProCysProAlaGlnTr 972  
2849 CGTCCGAGAGCGCGCGCTGTGAACAGAGTGCCTGCCCGCAGCAAGT 2898  
972 pArgLeuGlyAlaTrpSerGlnCysSerAlaThrCysGlyLysGlyIleG 989  
2899 GAAGAACAGACCTGTGAGTGTGAGTGTTCAGTACCTGCGGTGAAGAACCG 2948





```
595 tPheGluTyrGlnValCysAsnSerGluGluCysProGluYThrTyrGlnA 612
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1931 CTAGGATTTCCAGCTCTGCAACTGCGACGACTGCCCTGACGCGGTGGCCG 1980
612 sPpHeArGAlaGlnGlnCysAlaLysArgAsnSerTyrTyrValHisGln 628
      ||||||| :::::::::: :::::::::: :::::::::: |||
1981 ACTTCGCGAGAGACAGTGGCGGACGTGGACCTTCTGACGACATGCT 2030
629 AsnAlaLysHisSerTyrPValProTyrGluProAspAspAlaGlnLys 645
      :::::::::: :::::::::: :::::::::: ::::::::::
2031 GACGCTCAACACCACTGGCTGCCACGACGACCGGACCGCCAGAGAGCG 2080
645 sCysGluLeuLeuIecysGlnSerAlaAspThrGlyAspValAlaPheMet 662
      :::::::::: :::::::::: :::::::::: ::::::::::
2081 GTGTATCTCTACTGTAGTCCAGAGACGCGGAGGTGGTGTCCATGA 2130
662 snglValValHisAspGlyThrArgCysSerTyrArgAspProTyrSer 678
      :::::::::: :::::::::: :::::::::: ::::::::::
2131 AGCGATAGTGTCATGACGGGACACGCTGTCTCTCAAGGACGCTGACG 2180
679 ValCysAlaArgGlyGluCysValProValGlyCysAspLysGluValG 695
      :::::::::: :::::::::: :::::::::: ::::::::::
2181 CTCTGCTGCTGCTGGGACTGCGAGAGAGTGGCTGTGACGGGGTGTGG 2230
695 ySerMetLysAlaAspAspLysCysGlyValCysGlyGlyAspAsnSerH 712
      ||| ||| :::::::::: :::::::::: ::::::::::
2231 CTCACGACAGACAGGAGACAACTGTGTGTGTGGCGAGGGGACAACTCC 2280
712 IsCysArgThrValLysGlyThrLeuGlyLysAlaSerLysGlnAlaGly 728
      ||||| :::::::::: :::::::::: ::::::::::
2281 ACTGCAAGTGTGTCAAGGGCAGCTTCGCGCTGCCCAAGAACGTTGCT 2330
729 AlaLeuLysLeuValGlnLeuProAlaGlyAlaArgHisGlnGlnLeu 745
      :::::::::: :::::::::: :::::::::: ::::::::::
2331 TACATCAAGATGTTGAGATCCCGGACGAGACCACTGCTCAATCA 2380
745 uAlaLeuGluLysSerProHisArgSerValValLysAsnGlnValThrG 762
      :::::::::: :::::::::: :::::::::: ::::::::::
2381 GGAAGCAGACACACCAGCCATCACCCTGCGCTCAAAAACCTGAGACAG 2430
762 ySerPheIleLeuAsnProLysGlyLys...GluAlaThrSerArgThr 777
      ||||| :::::::::: :::::::::: ::::::::::
2431 GCAGGTTCATTTTAATGAGAGAAATGACGTGATCCCAACATCCCAAGCC 2480
778 PheThrAlaMetGlyLeuGluThrProLysAspAlaValGluAspAlaLys 794
      ||| ||||||| :::::::::: ::::::::::
2481 TTCATGCGCATGGCGGTGGAGTGGAGTACCGGGATGAGGACGGCGGGA 2530
794 uSerLeuLysThrSerGlyProLeuProGluAlaIleAlaIleLeuAlaL 811
      :::::::::: :::::::::: :::::::::: ::::::::::
2531 GACGCTGACAGACCATGGGCCCTCCACGACCATCACTGTGTGATCA 2580
811 euProProThrGluGlyGlyProArgSerSerLeuAlaLeuTyrLysTyrVal 827
      :::::::::: :::::::::: :::::::::: ::::::::::
2581 TCCCA.....GAGGGGAGCGCCGATCTCACTGACCTTCAAGATCAAG 2624
828 IleHisGlnAspLeuLeuProLeuIleGlySerAsnAsnValLeuLeuG 844
      ||||||| ||| :::::::::: ::::::::::
2625 ATCCATGAGGACCTGCTCAAT...GTGATGACACACAGCTCTCTGGAAGA 2671
844 uGluMetAspThrTyrGluThrAlaLeuLysSerTyrAlaProCysSerL 861
      :::::::::: :::::::::: :::::::::: ::::::::::
2672 CGACTCTGTGGCTATGAGTGGGCTTGAAGAAGTGTGCGCTCTGCTCA 2721
861 yAlaLysGlyGlyGlyIleGlnPheThrLysTyrGlyCysArgGlyArg 877
      ||| ||||||| ||| ||||||| ||| ||||||| |||
2722 AGCCCTGCGGTGGAGGTCCCAATTCACCAAGATGCTGCGCGGGAAG 2771
878 ArgAspHisMetValGlnArgHisLeuCysAspHisLysLysArgP 894
      ||||| ||||||| :::::::::: ::::::::::
2772 CTGAGCACCACAGATGGTGCACCGAGCTTCTGCACTCGTCCAAAGCC 2821
894 oLysProIleArgArgArgCysAsnGlnHisProCysSerGlnProValT 911
      :::::::::: :::::::::: :::::::::: ::::::::::

||||| ||||||| ||||||| :::::::::: ||||||| |||||||
2822 CAAGGCATCCCGGACCTGCAACCCACAGAGTGTCCAGCCCGTGT 2871
911 rPValThrGluGluThrProGlyAlaCysSerArgSerCysGlyLysLeuLys 927
      ||||||| ||||||| ||||||| ||||||| |||||||
2872 GGTCCAGGGGTAGTGGAGCCGTGCAGCGGAGCTGTGGGCGGACAGC 2921
928 ValGlnThrArgGlyIleGlnCysLeuLeuProLeuSerAsnGlyThrH 944
      :::::::::: ||| :::::::::: ::::::::::
2922 ATCCAGTTCGCTGTGTGGCTGTGTGTCAGCTTGCACACACACAGC 2971
944 sLysValMetProAlaLysAlaCysAlaLysArgProGluAlaArg 961
      :::::::::: :::::::::: :::::::::: ::::::::::
2972 CCGCTCGTGACACACCAACGACTGCATGAGCGCTGACCGGAGCGCG 3021
961 rGProCysLeuArgValProCysProAlaGlnThrArgLeuLysAlaTrp 977
      ||| ||| ||| ||| ||| ||| ||| ||| |||
3022 GGGCTGCACACCGGACGCTGTGCTGGCGGCTGGCGGCTGTGATCTGG 3071
978 SerGlnCysSerAlaThrCysGlyGluGlyIleGlnAlaArgGlnVal 994
      ||||||| :::::::::: :::::::::: ::::::::::
3072 TCCAGTGTCTAGTAACCTGTGGAACGACACCAAGAACGCGCACTGCT 3121
994 lCysArgThrAsnAlaAsnSerLeuGlyHisCysGlyGluLysArgProAl 1011
      :::::::::: :::::::::: :::::::::: ::::::::::
3122 CTGCGCAACTGCGGACGACAGTTCGGGTGTGCGGAGAGCGGCTGTG 3171
1011 sPThrValGlnValCysSerLeuProAlaCysGlyLysAsnHisGlnAsn 1027
      :::::::::: :::::::::: :::::::::: ::::::::::
3172 AGACGGACAGAGATCTGAGGCTTGGCCCTGTCCCGAACAACCTGTGAC 3221
1028 SerThrValArgAlaAspValTrpGluLeuGlyThrProGluGlyLnt 1044
      :::::::::: :::::::::: :::::::::: ::::::::::
3222 CCGTCCAAAGAGCTGCTGCTG.....GTCCAGTG 3250
1044 pValProGlnSerGlyProLeuHisProIleAsnLysIleSerSerMetC 1061
      ||||| :::::::::: :::::::::: ::::::::::
3251 GCTATCCCGACCGGACCCCAACTGCGCATGCCAGAGACCTGCTCA... 3296
1061 yAlaIleAlaGluProCysThrGlyAspArgSerValPheCysGlnMetGlu 1077
      ||| ||||||| ||| ||||||| ||| |||||||
3297 .....AAGGCGCGTGGCAAGGTGACAGTCAAGTCTGTAGATGAGA 3341
1078 ValLeuAspArgTyrCysSerIleProGlyTyrHisArgLeuLysCys 1094
      ||||||| :::::::::: :::::::::: ::::::::::
3342 GCTTGTCTGCTTATGCTCATCCAGGCTTACAAATGATGCTGTGCA 3391
1094 lSerCysIleLysLysAlaSerGlyProAsnProGlyProAspProGly 1111
      |||||
3392 GTCTGT..... 3398
1111 rOrhSerLeuProProPheSerThrProGlySerProLeuProGlyPro 1127
      :::::::::: :::::::::: ::::::::::
3399 .....AACCGGACGACAACTC 3416
1128 GlnAspProAlaAspAlaGluProPro...GlyLysProThrGlySer 1143
      ||| ||| ||| ||| ||| ||| ||| |||
3417 ACTGATGTGACAGAGGACGACCACTCTGTGGAGACCAATGACAT 3466
1143 rGluAspHisGlnHisGlyArgAlaThrGlnLeuProGlyAlaLeuAsp 1160
      ||| ::::::::::
3467 TGAAAG..... 3473
1160 hrSerSerProGlyThrGlnHisProPheAlaProGluThrProIlePro 1176
      :::::::::: ::::::::::
3474 .....CTCAGCCCGACCTTTCAGTCC 3497
1177 GlyAlaSerTrpSerIleSerProThrThrProGlyLeuLeuProG 1193
      :::::::::: ::::::::::
3498 ACTCTAGCATGAGGTGCAGCT.....CCGCGACGACATACC... 3536
1193 yTrpThrGlnThrProThrProValProGluAspLysGlnProGlyG 1210
      :::::::::: ::::::::::
```



730 GCGCTCTAGAGAGACAGCCACAGCTCGAGGCGGAGGCGACGACGCA 779  
253 SALALysProGlySerIleGluValIleuValAlaAspAsp 270  
780 TCGTCAGACATGATCACTACATCGAGTCTCTGGCGGTGATGACT 829  
270 eValValArgPheHisGlyLysGluHisValGlnAsnTyrValLeuThr 286  
830 CTGTGTGACAGTTCACGCGAAGAGCACGTACAGAACTCTCTGACA 879  
287 LeuMetAsnIleValAspGluIleTyrHisAspGluSerLeuGluValHis 303  
880 CTCATGAACATGTCAATGAATCTACCATGACAGATCTTGGGTGCCCA 929  
303 sIleAsnIleAlaLeuValArgLeuIleMetValGlyTyrArgInsrt 320  
930 CATCAACGTGGTCTGTCGCGATCATCTCTGAGTATGGAAGTCCA 979  
320 euSerLeuIleGluArgGlyAsnProSerArgSerLeuGluGlnValCys 336  
980 TGAGCCCTATCGAGATCGGAACCCCTCTCAGAGCTGAGAAATGTGTC 1029  
337 ArgTPrAlaHisSerGlnGlnArgGlnAspProSerHisAlaGluHisI 353  
1030 CGCTGGGCTTACTCTCAGCAGACAGACACAGGCGCCACATGAAATCCA 1079  
353 sAspHisValAlaPheLeuThrArgGlnAspPheGlyProSer..... 367  
1080 CGATCAGCGCATCTTCCTCACACGCGCAGAGACTTGGGCTTCGCGCATCG 1129  
368 ..GlyTyrAlaProValThrGlyMetCysHisProLeuArgSerCysAla 383  
1130 AAGGTATGCTCTCTCACCGGATGTCATCCGTCGCGACTGCGACCC 1179  
384 LeuAsnHisGluAspGlyPheSerSerAlaPheValIleAlaHisGluTh 400  
1180 CTGAACCATGAGAGAGCGGCTTCTCTCAGGCTTGTGTGGCCCATGAGAC 1229  
400 rGlyHisValLeuGlyMetGluHisAspGlyGlnLysnGlyCysAla 417  
1230 TGGCCACGTGCTGGGCATGAGCACGACGCGGAGGCAACCGCTGTGGCG 1279  
417 sPGLuThrSerLeuGlySerValMetAlaProLeuValGlnAlaIlePhe 433  
1280 ACGAGTGGCGGCTGGGCGACATCATGCGCCCTGTGTGAGGCGCCCTTC 1329  
434 HisArgPheHisTyrSerArgCysSerLysLeuGluLeuSerArgTyrLe 450  
1330 CACCGCTTCCACTGTGTCGCGTGCAGCCACGAGAGCTGAGCCGCTACT 1379  
450 uProSerTyrAspCysLeuLeuAspAspProPheAspProAlaTyrProG 467  
1380 GCACCTCTATGACTGCTCTGTGATGACCCCTTGGCCCGACGACTGGCGG 1429  
467 InProProGluLeuProGlyIleAsnTyrSerMetAspGluGlnCysArg 483  
1430 CGCTGCCCGAGCTCCGCGACTGCATCTCATGACGACGACGAATGCCGC 1479  
484 PheAspPheGlySerGlyTyrGlnThrCysLeuAlaPheArgThrPheG 500  
1480 TTTGACTTGGGCTGGGCTGATGATGTCACGCGCTTCGGGCTTTTGA 1529  
500 uProCysLysGlnLeuThrCysSerHisProAspAsnProTyrPheCysL 517  
1530 CCGCTGCAAGCAGCTGTGTGCGCATCTGACAAACCCCTACTTTTGA 1579  
517 ystThrLysLysGlyProProLeuAspGlyThrGluCysAlaProGlyLys 533  
1580 AGACCAAGAGAGGGGCCCTTGTGAGCGGACTATGTGTACACTGGCAAG 1629  
534 TrpCysPheLysGlyHisCysIleTyrLysSerProGluGlnThrTyrG 550  
1630 CATTTGTTTAAAGACATGCTGCTGACACTGAC...ATCTCAA 1676

550 yGlnAspGlyGlyTyrPheSerSerTyrThrLysPheGlySerSerArg 567  
1677 ACGGAGCGGACACTGGGCGCTTGAGTCCGTTGGCTCTCTACGTA 1726  
567 eCysGlyGlyGlyValArgSerArgSerArgSerCysAsnAsnProSer 583  
1727 CCGTGTGACAGGCGCTGAAGTTCAGGACCGCGCAGTGTGACAAACCCAC 1776  
584 ProAlaTyrGlyArgProCysLeuGlyProMetPheGluTyrGlnVal 600  
1777 CCGGCAACGGGCGCGCACTGCTGGCCCTTGGCTGACACTTCCAGCT 1826  
600 LysAsnSerGluGluCysProGlyTyrThrGluAspPheAlaGln 617  
1827 CTCAGCGCCGACAGACTGCCCGACCTCCCTGGCTGACTTCCCGAGGAC 1876  
617 InCysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHisSer 633  
1877 AGTCCGCGCAGTGGACCTGTACTTCCAGCAGCGCGACCCAGCAC 1926  
634 TrpValProTyrGluProAspAspAlaGlnLysCysGluLeuIleCys 650  
1927 TGGCTGCCCGACGAGCACCGGATGCCAAGAGAGATGCCACTGTACTG 1976  
650 sGlnSerAlaAspThrGlyAspValAlaPheMetAsnGlnValHisAla 667  
1977 CGAGTCCAGGAGACCGCGGAGGTGCTGTCCATGAAGCCGATGTGCATG 2026  
667 sPGLuThrArgCysSerTyrArgAspProTyrSerValCysAlaAlaGly 683  
2027 ATGGAGCGCGCTCTCTCACAGCGCCCTTCAAGCTGTGTGGCGGG 2076  
684 GluCysValProValGlyCysAspLysGluValGlySerMetLysAlaAs 700  
2077 GACTGCGAGAGAGTGGCTGTGACGCTGTATGCGCTCCAGCAAGAGAGA 2126  
700 PASPLysCysGlyValCysGlyLysAspAsnSerHisCysArgThrVal 717  
2127 AGACAAAGTGTGGCGTGTGCGAGGCGACAAACGCACTGCAAGGTGCTCA 2176  
717 ysgLThrLeuGlyLysAlaSerLysGlnAlaGlyAlaLeuLysLeuVal 733  
2177 AGGGCAGTTCACACGCTTACCCCAAGAGCATGTTTACATCAAGATTT 2226  
734 GlnIleProAlaGlyAlaArgHisIleGlnIleGluAlaLeuGluLysSe 750  
2227 GAGATCCCTGAGGAGCGACAGACACTGCTCATGAGGAGTAGACGCCAC 2276  
750 rProHisArgSerValValLysAsnGlnValThrGlySerPheIleLeu 767  
2277 CAGCCACCACTGGCGCTCAAGAACCTGAGACAGGCAAGTCAATCTTAA 2326  
767 snProLysGlyLys...GluAlaThrSerArgThrPheThrAlaMetGly 782  
2327 ATGAAGAAATGAGCTGATGCCAGTTCCAAACCTTCATTCGATGGCG 2376  
783 LeuGluThrGluAspAlaValGluAspAlaLysGluSerLeuLysThrSe 799  
2377 GTGAGTGTGGAGTACAGAGACAGAGACGCGCGGAGACGCTGAGACAT 2426  
799 rGlyProLeuProGluAlaIleAlaIleLeuAlaLeuProProThrGlu 816  
2427 GGGCGCCCTCCAGCGCACATCACCGTTCGTGTATGTCG....GTGG 2470  
816 LysGlyProArgSerSerLeuAlaTyrLysTyrValIleHisGluAspLeu 832  
2471 GAGACACCGCGGTCTCACTGACGTACAAATACATGATTCATGAGAGACTCA 2520  
833 LeuProLeuIleGlySerAsnAsnValLeuLeuGluGluMetAspThrTyr 849  
2521 CTGAAT...GTGATGACACACAGTCTCTGGAAGAGAGACTGTGTGTCTA 2567

```

849  rGlUTPAlaLeuLysSerTrpAlaProCysSerLysAlaCysGlyGly 866
      |||||
2568  CGAGTGGGCGCCGTGAAGAAGTGTCTCCGTCTCAAGCCCTGGGCGAG 2617
      |||||
866  LylleGlnPheThrLysTyrGlyCysArgArgArgArgAspHisMet 882
      |||||
2618  GGTCGCCAGTTCACCAAGATATGCTGCCGCGAGGCTGGACCAAGAAG 2667
      |||||
883  ValGlnArgHisLeuCysAspHisLysLysArgProLysProLleArg 899
      |||||
2668  GTACACCGCTGGCTTGTGTGGCCGCTCTCTCAACCCCAAGCATCGG 2717
      |||||
899  gATgCysAsnGlnHisProCysSerGlnProValTrpValThrGln 916
      |||||
2718  AGCGTGCACCCAGAGAAATGCTCCAGCAGGTGGTGCACAGCGAAT 2767
      |||||
916  rpgLysAlaCysSerArgSerCysGlyLysLeuGlyValGlnThrArg 932
      |||||
2768  GGGAGCCATGTAGCCAGACCTGTGGCGGACAGCATGCGTGGCTCC 2817
      |||||
933  lLeGlnCysLeuLeuProLeuSerAsnGlyThrHisLysValMetPro 949
      |||||
2818  GTGCGCTGATTCAGCCGCTACAGCAGACACCAGCCGCTCGTCCAGC 2867
      |||||
949  aLysAlaCysAlaGlyAspArgProGlnAlaArgArgProCysLeuArg 966
      |||||
2868  CAAGCACTGCATAGACGCCGCCGAGAGCGCGGCTCGACCCGCG 2917
      |||||
966  aLProCysProAlaGlnTrpArgLeuGlyAlaTrpSerGlnCysSer 982
      |||||
2918  AGCTGCGCCCTGCTGCTTGGCAGCGGCGCTGCTGCCAGTGCACAT 999
      |||||
983  ThrCysGlyGlyGlyTleGlnGlnArgLysValCysArgThrAspAl 999
      |||||
2968  ACCGTGGCCAGCAGCACCAGAGCGGCGCATGCTGCCAGCAGCGG 3017
      |||||
999  aAsnSerLeuGlyHisCysGlyGlyAspArgProAspThrValGln 1016
      |||||
3018  CCACAGCTTCGGCATCTCCAGAGAGCGTCTGAGACAGCAGAGACT 3067
      |||||
1016  ySerLeuProAlaCysGlyGlyAsnHisGlnAsnSerThrValArgAl 1032
      |||||
3068  GCAGGCTTGGCCCTGCTCCGGAACATCTCAGATCCCTCCAGAGAAG 3117
      |||||
1033  AspValITrpgLysLeuGlyThrProGlnGlyLysTrpValProGln 1049
      |||||
3118  TACGTA.....CTTCAGTGGCTGTCCGCCCGGA 3146
      |||||
1049  yProLeuHisProIleAsnLysLleSerSerMetCysAlaLleGln 1066
      |||||
3147  CCCCAGCTCGCCCATCCGGAAGATCTGCTCA.....AAGGCCACT 3187
      |||||
1066  ySThGlyAspArgSerValPheCysGlnMetGluValLeuAspArgTyr 1082
      |||||
3188  GCCAAGCGCACAAGTATATTCTGTAGAGTGAAGTCTTCTCCGCTAT 3237
      |||||
1083  CysSerIleProGlyTyrHisArgLeuCysValSerCys..... 1096
      |||||
3238  TGCTCATTCACAGGCTACACACACCTGCTCTCAAGTCTGTAACTGT 3287
      |||||
1097  .....lLeLysLysAlaSerGlyProAspProGlyPro. 1107
      |||||
3288  CACACACCTCACCAAGTGGAGGCGACAGATAGAGCCAGCCCTGGGAG 3337
      |||||
1108  .....AspProGlyProThrSerLeuProProPheSerThrPro 1120
      |||||
3338  ACAACGACATTCAGCTGTATGCTTACCTCCAGTGGCCACTGTAGGC 3387
      |||||
1121  GlySerProLeuProGlyProGlnAspProAlaAlaLleLysProP 1137
      |||||
3388  ATGGAGGGGGGGCATACCAAGACCCCTGGAGGTCCCTCTCAATGC 3437
      |||||
1137  oGlyLysProThrLysSerGlnAspHisGlnHisGlyArgAlaThrG 1154

```

```

seq_name: /SID5/gcgcdata/geneseq/geneseqn-emb1/MA1998.DAT:AAV06592
seq_documentation_block:
ID  AAV06592 standard; cDNA; 6692 BP.
XX
AC  AAV06592;
XX
XX  06-JUL-1998 (first entry)
XX
DE  Human N-proteinase (long form) cDNA.
XX
KW  N-proteinase; human; collagen; antibody; rheumatoid arthritis;
KW  fibrosis; Ehlers-Danlos disease; diagnosis; gene therapy; ss.
XX
OS  Homo sapiens.
XX
FH  Key
FT  Location/Qualifiers
FT  CDS 82..371
FT  CDS /*tag= a
XX
XX  WO9800555-A1.
XX
PD  08-JAN-1998.
XX
PF  03-JUL-1997; 97MO-US12427.
XX
PR  02-JUL-1997; 97US-0886333.
PR  03-JUL-1996; 96US-0021203.
XX
XX  (COL1) COL1GE A.
XX  (LAP1) LAP1ERE C.
XX  (PROC) PROCKOP D J.
XX
XX  Col1ge A, Lap1ere C, Prockop DJ;
XX
XX  WPI; 1998-086980/08.
XX  P-PSDB; AAW47028.
XX
PT  Polynucleotide sequence encoding human N-proteinase - used to
PT  produce mature collagen in vitro and antibodies to treat fibrosis
PT  and rheumatoid arthritis
XX
PS  Claim 1; Fig 1A-C; 49p; English.
XX
XX  This nucleic acid sequence includes a coding region for the 130 kDa
XX  long form (see AAW47028) of human N-proteinase, the enzyme
XX  responsible for cleaving N-propeptide from procollagen to produce
XX  mature collagen. Overlapping cDNA clones were isolated from a
XX  human skin fibroblast cDNA library using bovine N-proteinase cDNA
XX  (see AAV06594) as probe, and used to generate the full-length cDNA.
XX  A nucleic acid sequence (see AAV06593) for the 70 kDa short form (see
XX  AAW47029) of human N-proteinase was also generated. The isolated
XX  nucleic acids can be utilized in the recombinant production of
XX  human N-proteinases in claimed transformed or transfected host cells.
XX  They can also be used diagnostically for detection of N-proteinase
XX  expression and in gene therapy approaches to remedy deficiencies
XX  of functional N-proteinase in various disease states such as
XX  Ehlers-Danlos disease. The N-proteinase may be used for the
XX  production of mature collagen in vitro and for the production of
XX  antibodies which may be used for diagnosis and therapy of diseases
XX  including fibrosis and rheumatoid arthritis.

```



238	gLyLeuValGlyaspGlnLeuGlyaspPrlnGlnAlaGlys...	ArgArgHn	253
811	GGCGTCTAGAGGAGCGACCGCCAAACAGCTCCAGGCGGAGGCGACGCGAGGCA		860
253	sAlaLysProGlySerTyrSerIleGluValLeuLeuValAlaPaspAsp		270
861	TGCTGGCAGACGATGACTCAACAACATCGAGGTCCTGGCTGGCGGTGAGTACT		910
270	exValValArgPheHisGlyLysGlnHisValGlnAsnTyrValLeuThr		286
911	CTGTGGTGCAGATTCCACGGGAGAGGACGACTACAGAAATGCTGGCTGACA		960
287	LeuMetAsnIleValAspGlnIleTyrHisAspGlnSerLeuGluValIleI		303
961	CTCATGACATCTTGCTCAATGAAATCTACCATGACGACGATCTTTGGGTGCCA		1010
303	sIleAsnIleAlaLeuValAlaArgLeuIleMetValGlyTyrArgGlnSerI		320
1011	CATCAACGCTGGCTCTGGTGGCGGATCATCTCTCCAGCTATGGAAGATCCA		1060
320	eusSerIleIleGlnAlaArgLysAsnProSerArgSerLeuGlnValAlcys		336
1061	TGAGCTTCATGAGATCGGGAAACCCCTCTCCAGACCTGGAAATGCTGCG		1110
337	ArgTPrAlaHisSerGlnGlnAlaArgLysAspProSerHisAlaGlnHisH		353
1111	CGCTGGGCTCACCTCCACACAGACGACGACCGGGCCACGATGAAATACA		1160
353	sAspHisValValPheLeuThrArgGlnAspPheGlyProSer.....		367
1161	CGATACGCGCATCTCTCCACACGGGACGAGACTTTGGGCTTCCGGCATGC		1210
368	..GlyTyrAlaProValThrGlyMetCysHisProLeuArgSerCysAla		383
1211	AAGGTTATGCTCTCTCCACCGGATGTCGCAATCGCGTCCGACGTGACCC		1260
384	LeuAsnHisGlnAspGlyPheSerSerAlaPheValIleAlaHisGlnThr		400
1261	CTGAACCATGAGAGCGGCTCTCCACGCTTTGTGGTGGCCCATGAGAC		1310
400	rgLysHisValLeuGlyMetGlnHisAspGlyGlnGlyAsnGlyCysAlaA		417
1311	TGGCCACGTCGTGGGCGATGAGACACGAGCGGCGAGGGCAACGCTGTGGC		1360
417	spGlnThrSerLeuGlySerValMetAlaProLeuValGlnAlaPhe		433
1361	ACGAGGATGGGCTGGGACACATATGGCGCCCTGTGACGAGCCGCTTC		1410
434	HisArgPheHisTyrSerArgCysSerTyrLeuGlnLeuSerArgTyrLe		450
1411	CACGGCTTCCACTGGTCCGCTGACGCCAGCGAGGATGACGCCCTACTCT		1460
450	uproSerTyrAspCysLeuLeuAspAspProPheAspProAlaTyrProG		467
1461	GCACCTCTTATGACTGCTCTCGATGATGACCCCTTGGCCACGACTGGCCG		1510
467	lnproGlnLeuProGlyIleAsnTyrSerMetAspGlnIleCysArg		483
1511	CGCTGCCACAGCTCCGGGACATCGAATCTCATGAAACAGAACGAAATGCCG		1560
484	PheAspPheGlySerGlyTyrGlnThrCysLeuAlaPheArgThrPheG		500
1561	TTTGACTCTGGCGCTGGGCTATGATGATGACACGGCGCTTCCGACCTTTGA		1610
500	uproCysLysGlnLeuThrCysSerHisProAspAsnProTyrPheCysL		517
1611	CCGCTCCAAACACTGTGGGACGACCATCTGACAAACCCCTTACTTTTGA		1660
517	ysThrLysLysGlyProProLeuAspGlyThrGlnCysAlaProGlyLys		533
1661	AGACCAAGAAAGGGGCCCCCTTGGAGCGGAGCTATGTGTGCACCTGGCAAG		1710
534	trpCysPheLeuGlnHisCysIleTyrPlysSerProGlnGlnThrTyrG		550



```

1711 CATGTTTAAAGACACTGCTCATCTGCGACACCTGAC...ATCCTCA 1757
550 YGlnAspGlyGlyTyrSerSerTrpThrLysPheGlySerCysSerArgS 567
1758 ACCGGACGGAGCTGGGGCTTGAGTCCGTTTGGCTCTCTGCTCAACGTA 1807
567 eRCysGlyGlyValArgSerArgSerArgSerCysAsnAsnProSer 583
1808 CCGTGGCAGGGCGGTGAAGTTAGAACCCCGCAGTGTACAAACCCACAC 1857
584 ProAlaTyrGlyValArgProCysLeuGlyProMetLysPheGluTyrGlnVa 600
1858 CCGGACCAACCGGGGCGCGACCTGCGGCGCTTGCGTACGACTTCCAGCT 1907
600 LysAsnSerGluGluCysProGlyThrTyrGlnAspPheAlaGlnG 617
1908 CTCGACGCCCGACGACTGCGCGACCTGCGTCCCTGCTGACTTCCGCGAGAGC 1957
617 IncysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHisSer 633
1958 AGTGGCGCCAGTGGGACCTGTACTTCCAGACAGCGCGACCCAGACACAC 2007
634 TrpValProTyrGluProAspAspAlaGlnLysCysGluLeuIleCys 650
2008 TGCCTGCCCCACGAGACCGCGGATGCCAAGAGAGATGCCACCTGACTG 2057
650 sGlnSerAlaAspThrGlyAspValValPheMetAsnGlnValValHisA 667
2058 CAGGTCACGAGGAGACCGGGAGGTGTGTCCTCAAGCGCATGTGGCATG 2107
667 spGlyThrArgCysSerTyrArgAspProTyrSerValCysAlaArgGly 683
2108 ATGGGACGGCTGTCTCTACAGAGAGCGCTTCAGCCTGTGTGCGCGGG 2157
684 GluCysValProValGlyLysAspLysGluValGlySerMetLysAlaAs 700
2158 GACTGAGAGAGGTGGCGGTGACGCTGATCGCTCCAGCAAGCAGAGA 2207
700 pAspLysCysGlyValCysGlyLysAspAsnSerHisCysArgThrAla 717
2208 AGACAAAGTGTGGCGTGTGGGAGGAGACACACCCACTGCAAAATGGTGA 2257
717 ySGlyThrLeuGlyLysAlaSerLysGlnAlaGlyAlaLeuLysLeuVal 733
2238 AGGGCAGCTTCACACGCTACCCCAAGCATGGTGTACATCAAGATGTTT 2307
734 GlnIleProAlaGlyAlaArgHisIleGlnIleGlnAlaLeuGlnLysSe 750
2308 GAGATCCCTCGACAGACGCCAGACACCTGCTCATTCAGAGAGTAGACGCC 2357
750 TrpHisArgSerValValLysAsnGlnValThrGlySerPheIleLeuA 767
2358 CAGCCACCATCTCGCCGTCAGAGACCTGGAGACAGGCAAGTTCACTTA 2407
767 snProLysGlyLys...GlnAlaThrSerArgThrPheMetAlaMetGly 782
2408 ATGAAGAAATGACGTGATGTCAGTTCCAAACCTTCATTTGCCATGGCG 2457
783 LeuGluTrpGluAspAlaValGluAspAlaLysGluSerLeuLysThrSe 799
2458 GTGGAGTGGGAGTACAGAGACGAGAGCGCGGAGAGCGTGCAGACCAT 2507
799 rGlyProLeuProGluAlaIleAlaIleLeuAlaLeuProProThrGlu 816
2508 GGGGCCCTCCACGGCACCATCACCGTTGTGTCATCCG.....GTGG 2551
816 LysGlyProArgSerSerLeuAlaTyrLysTyrValIleHisGluAspLeu 832
2552 GAGACACCGGGGTCTACGTACGTACAAATACATGATCCATGGGAGCTCA 2601
833 LeuProLeuIleGlySerAsnAsnValLeuLeuGluGluMetAspThrTy 849

```

---

```

2602 CTGAAT...GTGATGACAACAACGTCCTGGAAGAGACTGTGTGATTA 2648
849 rGluTrpAlaLeuLysSerTrpAlaProCysSerLysAlaCysGlyGlyG 866
2649 CAGTGGGCGCCCTCAGAAAGTGTCTGCTGCTCCAGAGCCCTGTGGGGAG 2698
866 LysIleGlnPheThrLysTyrGlyCysArgArgArgArgAspHisMet 882
2699 GTTCCAGCTACCAAGTATGCTGCGCGGAGGCTGGAGCACAAGATG 2748
883 ValGlnArgHisLeuCysAspHisLysLysArgProLysProIleArgAr 899
2749 GTACACGGTGGCTTCTGTCCCGCCCTCTCGAAGCCCAAGCATCCGAG 2798
899 gArgCysAsnGlnHisProCysSerGlnProValTrpValThrGluTr 916
2799 AGCGTCACAACCCACAGAGATGCTCCAGCAGCATGTGGGTACAGCGMAT 2848
916 TrpValLysSerSerArgSerCysGlyLysLeuGlyValGlnThrArgGly 932
2849 GGGAGCCATGTAGCCAGACCTGTGGCGGACAGCATGCAGGTGGCTCC 2898
933 IleGlnCysLeuLeuProLeuSerAsnGlyThrHisLysValMetProAl 949
2899 GTGCGCTGCATTCAGCGCTTACAGCAGACACACACCCTCCGTCGACAGC 2948
949 aLysAlaCysAlaGlyAspArgProGluAlaArgArgProCysLeuArgV 966
2949 CAAGCACTCATGACGCCCGCGCGAGAGCGCGCGCGCGCGCGCGCGCG 2998
966 aLProCysProAlaGlnTrpArgLeuGlyAlaTrpSerGlnCysSerAla 982
2999 ACCTGTGCGCTGTGTGTGGAGCGCGCGCGCTGTCTCCAGTCTCAGTA 3048
983 ThrCysGlyGluGlyIleGlnGlnArgGlnValValCysArgTrpAsnAl 999
3049 ACCTGTGGAGAGCGACCCAGAGAGCGCACTGTCTGCGACCGCGGGA 3098
999 aAsnSerLeuGlyHisCysGlyLysAspArgProAspThrValGlnValC 1016
3099 CGACAGCTTCGCGCATCTCCAGAGAGCGCTGAGACAGCGAGGAGACT 3148
1016 ySerLeuProAlaCysGlyLysAsnHisGlnAsnSerThrValArgAla 1032
3149 CGAGCTTGCGCCCTGTGTCGCCGAACATCTTCAGATCCCTCCAGAAAGC 3198
1033 AspValTrpGluLeuGlyThrProGluGlyGlnTrpValProGlnSerG 1049
3199 TACGTA.....GTTCAGTGGCTGTGCCCGCGCGGA 3227
1049 yProLeuHisProIleAsnLysIleSerSerMetCysAlaAlaGluProC 1066
3228 CCGGACCTCGCCATCCGGAAGATCTGTCA.....AAGGCCACT 3268
1066 yThrGlyAspArgSerValPheCysGlnMetGluValLeuAspArgTyr 1082
3269 GCCAAGGGGAGCAATCAATATCTGTAGATGGAAGTCTTGTGCCCTAT 3318
1083 CysSerIleProGlyTyrHisArgLeuCysCysValSerCys..... 1096
3319 TGTTCATCTCCAGGCTACCAAGCTGTGCTCAAGTCTGTAACTGTA 3368
1097 .....IleLysLysAlaSerGlyProAsnProGlyPro. 1107
3369 CAACAACCTCACCAAGCTGAGGGCAGATAGACCAACCGCTGGGAAGC 3418
1108 .....AspProGlyProThrSerLeuProProPheSerThrPro 1120
3419 ACAACGACATGAGCGTTCATGCTACCTCCACAGTGCACACTGAGGC 3468
1121 GlySerProLeuProGlyProGlnAspProAlaAspAlaAlaGluPro 1137
3469 ATGAGGTGGCGCATCCACCAAGACCCCTGGAGGTCTCTCAATGC 3518

```



[illegible]



Thu Aug 29 10:45:52 2002

us-09-938-330-20.rng

Page 27





```

102 TGGCCGGGATGCTACTCCATGATGACCAATGCCGGTTGCACCTTGGC 151
488 SerGlyThrGlnThrCysLeuAlaPheArgThrPheGluProCysLysG1 504
152 ACTGGGTACACACCTGCTTACGTTCCAGACCTTGGACCTTGCAGAGA 201
504 nLeuTrpCysSerHisProAspAsnProTyrPheCysLysThrLysLysG 521
202 GCTATGATGGAGCCACCCCTGACAAACCTTCTTCCAGACCAAGAAAG 251
521 LyrProLeuAspGlyThrGluCysAlaProGlyLysTrpCysPheLys 537
252 GGGCGGGCCTGGAGTGGACAGAGTGGCCAGCAGAGTGGTCTTCAA 300
538 G1HisCysL1eTrpLysSerProGluGlnThrTyrGlyLysAspGlyG1 554
301 GGCCA.TGTATCTGGAAGTCCACAGAGCAAACTTATGCCAGATGAGAG 349
554 YTPSerSerTrpThrLysPheGlySerCysSerArgSerCysGlyGlyG1 571
350 CTGGAGTCTCTGGACCAA.TTTGGCTCATGTTCTCGTGGTGGAGAG 398
571 LysAlaGSerArgSerArgSerCysAsnAsnProSerProAlaTyrGly 587
399 GGGTGCATCCCGAAGCCGAGAGCTGGAGCAACCTCTCCAGCTATGGA 448
588 G1AArgProCysLeuGlyProMetPheGluTyrGlnValLysAsnSerG1 604
449 GGCGGCGGCTGCTCAGAGGCTCATGTTGAGTACCAGATGCAACAGTGA 498
604 uGluCysProGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaLys 621
499 GGAAGTCTCTGGGCGCTTACAGAGACTTCCAGCCAGACAGTGTGCCAAG 548
621 rGAsnSerTyrTyrValHisGlnAsnAlaLysHisSerTrpValProTyr 637
549 GAACTCTACTATATCCACAGATGCTCAAGCAGAG.TGGTGGCTATG 597
638 GluProAspAspAlaGlnLysCysGlnLeuLeuLecysGlnSerAlaAs 654
598 AGCCCGAGAGGTGATGCCAGAAAGTGCAGCTCATTTCA...GTCTCGA 644
654 PThGlyAspValAlaPheMetAsnGlnValValHisAspGlyThrArgC 671
645 CACTGGCATATGCTGCTCATGAAACAGTTGCC.CAACCAAGGAGACAGCT 693
671 ySerTyrArgAspProTyrSerValCysAlaArgGlyGluCysValPro 687
694 GAGTATCGCGAAG...TTAAAGTTTGGGCC...GGGAGTGTTCCT 737
688 ValGlyCysAspLysGluValGlySerMetLysAlaAspAspLysCysG1 704
738 .....GG 739
704 yValCysGlyGlyAspAsnSerHisCysArgThrValLysGlyThrLeu. 720
740 TGGCGCGGGGGGGCC.....CGCAGCGTGCAGAGGGGGGCTCC 777
721 Gly.....LysAlaSerLysGlnAlaGlyAlaLeuLysLeuVa 733
778 GGGGACACCCCGCACCGGGCGGGCGGGGGGGTCT..... 816
733 LglnLeuProAlaGlyAlaArgHisL1eGlnL1eGlnAlaLeuGluLys 750
817 .....CGGGCGGGGGCGGGCGGGCGGGAGAA..... 840
750 erProHis 752
841 ..CCCCAC 846
seq_name: gb_hnc:BC003269
seq_documentation_block:

```

```

LOCUS      BC003269                2931 bp      mRNA           linear      HTC 12-JUN-2001
DEFINITION Mus musculus, similar to a disintegrin-like and metalloprotease
            (repolysin type) with thrombospondin type 1 motif, 1, clone
            IMAGE:349206, mRNA.
ACCESSION  BC003269
VERSION    BC003269.1  GI:14707747
KEYWORDS   HTC.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 2931)
AUTHORS   Strausberg, R.
TITLE      Direct Submission
JOURNAL    Submitted (20-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
            Email: gcaps-r@mail.nih.gov
            Tissue procurement: Jeffrey Green M.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villaloboscm.tmc.edu.
            Vallatón, D.K., Luna, R.A., Hale, S.M., Huliy, S., Lu, X., Garcia,
            A.M., Hollway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
            Muzny, D.M., Gibbs, R.A.
FEATURES   source
            1..2931
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:349206"
            /tissue_type="Mammary tumor, C3(1)-tag model. Infiltrating
            ductal carcinoma. 5 month old virgin mouse."
            /clone_id="NCI-CGAP_Mam6"
            /lab_host="DH10B"
            /note="Vector: pCMV-Sport6"
BASE COUNT  708 a      812 c      811 g      600 t
ORIGIN
alignment_scores:
    Quality: 953.50      Length: 913
    Ratio: 1.884        Gaps: 35
    Percent Similarity: 55.422      Percent Identity: 29.025
alignment_block:
US-09-938-330-20 x BC003269 ..
Align seg 1/1 to: BC003269 from: 1 to: 2931
3 ProLeuArgAlaLeuLeuSerTyrLeuLeuProLeuHisCysAlaLeuGly 19
111 :::: 111 :::: 111 :::: 111 ::::
322 CCAGCGAAGCTTACGCGTCTAAACGTATATCCAGACCAAGAGTTCTTG 371
19 strAla...AlaGlySerArgThrProGluLeuHisLysSerGlyLys. 34
111 :::: 111 :::: 111 :::: 111 :::: 111 ::::
372 CGGTGCGAGGTGGGTGCGTCCATGCCAAGAGTCCCTTGGGGTCAAGCA 421
35 .....LeuSerAspTyrGlyValThrValProCysSerThrAsp 47
11111 111 ::::

```



```

422 AGCAGAACCCCTGCTCCGACATGGG...GACGTCCAGCGGGCAGCAGAA 468
48 PhearglyIyArpheLeuSerHisValValSerGlyProAlaAlaLase 64
469 TCTGGGGGCTCTGTGCGCACATGCTGTG..... 501
64 rAlaGlySerMetValValAspThrProThrLeuProArgHisSers 81
502 .....CTGCTCTGCTTCCATACATGCTGCTATGTGGCGGGGGG 544
81 erHisLeuArg.....Val 85
545 CACACAGCGGGCCCGCAGAGAAATGAGAGCTGATGCTGCTGCTGCTG 594
86 AlaArgSerProLeuHisProGlyIyThrLeuTyrProGlyArgValGly 102
595 GAGCGGGGGCGCGCAGATTCACACC..... 624
102 YArGHisSerLeuTyrPheAsnValThrValPheGlyLysGluLeuHisL 119
625 .....ACAGCCTTCGTCTGAGCGCTTGGCCAGCAGCTACATC 664
119 euArGLeuArgProAsnArgArgLeuValValProGlySerSerValGly 135
665 TGAAGTTGGAGCGGACAGCGGTTCTTGGCGCTTGCTTCACCTGCAG 714
136 TrpGlnGluAspPheArgGluLeuPheArgGlnProLeuArgGln..... 150
715 .....ACTGTGGGGCGCAGTCCCGGGTCCGAGGCACA 746
151 .....GluCysValTyrThrIyGly 158
747 ACATGTGACCCCGGAGGACCTGCTACCTTCTACTCTGCGACGG 796
158 aThrGlyeProGlyAlaAlaValAlaIleSerAsnCysAspGlyLeu 174
797 TGAAGGTCATCCCGCTCTGCGGACGCTCAGCTCTGTGAAGGTGTG 846
175 AlaGlyLeuIleArgThrAspSerThrAspPheHeIleGluPro..... 189
847 CGTGTGCTTCTTACCTACAAAGAGAGATTTTCATTCAGCAGGGCC 896
190 .....LeuGluArgGlyGlnGlnGluLysGluAs 200
897 TGGAGTGGCCACCGAGCGCTGGCCCTCGCTGCCGAGGAGAGATCAT 946
200 erGlyArgThr.....HisValValTyrArgArgGlu..... 210
947 CCGCAGCGCCCGCAGTTCACATCTGAGAGCGAAGCGGCGGCGCATGGC 996
211 .....AlaValGlnGlnGluTyrPalaGluProAspGlyAspLe 223
997 GCGCGCAAGTGGCGGTCATGAGCAGAGACCTGCCAACACAGCGATC 1046
223 uHisAsnGluAla..... 227
1047 GCGACCGGAGGCCAAGAACCCGGAACCACTGGCTGTCGGGAGCCCA 1096
228 .....PheGlyLeuGlyAspLeuProAsnLeu 236
1097 CCGCTCAGGAGCGGGAAGCCATCAGCAGCAGAAAGCATA..... 1137
237 LeuGlyLeuValGlyAspGlnLeuGlyAspThrGluArgIyAspArgH 253
1138 .....AGGAAGAGGAGATT 1151
253 sAlaLysProGlySerTyrSerIleGluValLeuLeuValValAspAs 270
1152 TGTGTCCAGCCCCCGTTAT...GTGGAACCATGCTGTGGCTGACCATG 1198
270 erValValArgPheHisGlyLysGlnHisValGlnAsnTyrValLeuThr 286
1199 CCATGGCGGACTTCCACGGC...AGCGGTCTAAAGCATTAACCTTTAAC 1245

```

---

```

287 LeuMetAsnIleValAspGluIleTyrHisAspGluSerLeuGlyValHi 303
1246 CTGTCTCGGTGGCAGCCAGGTTTTCAGCATCCAGCATTAAGAAATTC 1295
303 sIleAsnIleAlaLeuValArgLeuIleMetValGlyTyrArgGlnSerL 320
1296 AATTAGCTGTGTGTGTAAGATCTTGTGATA...TATGAGGAGCAGA 1342
320 euSerLeuIleGluArgGlyAsnProSerArgSerLeuGlnValLys 336
1343 AGGACCCAGATTAACCTCAATGACGCTTCACCTTCGGAATTTCTGC 1392
337 ArgTrpAlaHisSerGlnGlnArgGlnAspProSer.....HisAlaG 351
1393 AACTGG.....CAGAAACACACAAAGCCCCATGTGACCGGATCCAGA 1436
351 uHisHisAspHisValValPheLeuThrArgGlnAspPheGlyProSer. 367
1437 GCACATGACACTGCACATCTGTTCCACAGACGATTTATGTGGCTCC 1486
368 .....GlyTyrAlaProValThrGlyMetCysHisPro 378
1487 ACACGTGTACACTCTCGGATGGCAGATTTGGAACTGTATGTGACCC 1536
379 LeuArgSerCysAlaLeuAsnHisGluAspGlyPheSerAlaPheVa 395
1537 ACAGAGAGCTGCTCAGTATAGATGATGTTGCAAGCGGCTTCAC 1586
395 lIleAlaHisGluThrGlyHisValLeuGlyMetGluHisAspGlyGln 412
1587 CACAGCCCAAGATTTGGCCAGTGTTCATACATGCCGACAT...GATG 1633
412 lYAsnGlyCysAlaAsp.....GluThrSerLeuGlySerValMet 425
1634 CTAAACACTGTGCCAGCTTGAATGATGACTGGCGATTCATCATGTATG 1683
426 AlaProLeuValGlnAlaAlaPheHisArgPheHisTrpSerArgCys 442
1684 GCTTGATGCTCTCCAGCTTAGACCATAGCCAGCCCTGCTCAGCTTCC 1733
442 rLysLeuGluLeuSerArgTyrLeuProSer.....TyrAspCysLeu 457
1734 TGCCATACATGGTCACAGTCTTCTAGATATAGCAGCGGGAATGTGGA 1783
457 euAspAspProPheAspProAlaThrProGlnProProGluLeuProGly 473
1784 TGCACAAAGCCCAAGATCCATC...AAGTCCCTTGTGATCTTCCCGGT 1830
474 lLeAsnTyrSerMetAspGlnGlnCysArgPheAspPheGlySerGly 490
1831 ACCTTGATGATGCCAACCGCAGTGTACATTCATTCGGAAGAAATC 1880
490 rGlnThrCysLeuAlaPheArgThrPheGluProCysLysGlnLeuTrp 507
1881 CAAGCATGCGCT.....GATGAGCCACACATGATACCTCTGTGT 1924
507 ysSerHisProAspAspProTyrPhe...CysLysThrLysLysGlyPro 522
1925 GCACGTGCACCTCGGTGCTTACTGTGTGCAAAACAAACCTTCCCT 1974
523 ProLeuAspGlyThrGlyCysAlaPheGlyLysTrpCysPheLysGlyH 539
1975 TGGCAGATGCGACACCTGTGGAGAAGGAATGTGTGTCTGAGTCCAA 2024
539 sCyslIleTrpLysSerProGluGlnThrTyrGlyGln.....AspGlyG 554
2025 GTGGGTGAAGAACAGACATGAAGCATTTTCTACTCTGTTCATGAA 2074
554 lYTrpSerSerTrpThrLysPheGlySerCysSerArgSerGlyGly 570
2075 GCTGGGACCATGAGGACCTGGGAGACTGTCAAGAACCTGTGTGTGT 2124

```



```

|||||
373 AAATCACCAGAACTCCACGCTGAGGGCCGATCTCGGAACTTGGGACCC 324
1040 roglu[glu]in[trp]val[pro]glu[ser]gly[pro]leu[his]pro[ile]asn[lys] 1056
323 CAGAGGGGAGTGGGTGGCCACAAATCTGAGCCCTACATCCCATTAACAG 274
1057 l[le]ser[met]cys[ala]ala[glu]pro[cys]thr[glu]asp[arg]ser[val]ph 1073
273 ATATCATCA.....ACGAGCCCTGACAGGGAGACAGTCTGCTT 233
1073 eCys[glu]met[glu]val[leu]asp[arg]tyr[cys]ser[ile]pro[glu]tyr[his]a 1090
232 CTGCACAGATGGAAGTCTCGATCGTACTGCTCATTCGCCGCTACACACC 183
1090 r[glu]cys[glu]val[ser]cys[ile]lys[lys]ala[ser]gly[pro]asn[pro]gly 1106
182 GCGTCTGCTGTGTGTCTGTCTGTCAAGAAAGGCGGCGCCCAACCCCTGGC 133
1107 Pro[asp]pro[glu]tyr[thr]ser[leu]pro[pro]phe[ser] 1118
132 CCAGACACTGGCCCACTCACTGCTCTCTCTCG 97
seq_name: gb_est1:BI081935

```

```

seq_documentation_block:
LOCUS BI081935 724 bp mRNA linear EST 20-JUN-2001
DEFINITION 602877161F1 NC1_GAP_Mam2 Mus musculus cDNA clone IMAGE:5008982 5',
ACCESSION BI081935 mRNA sequence.
VERSION BI081935.1 GI:14500265
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```

```

REFERENCE NIH-MGC http://mgc.ncl.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cga@bbs-rmail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: L14M11054 row: 1 column: 15
High quality sequence start: 2
High quality sequence stop: 705.
Location/Qualifiers
1..724

```

```

FEATURES
source

```

```

/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5008982"
/clone_id="NC1_GAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; site_1: SalI;
site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 162 a 208 c 217 g 137 t
ORIGIN

```

```

alignment_scores:
Quality: 860.50 Length: 243
Ratio: 4.413 Gaps: 2

```

```

Percent Similarity: 80.247 Percent Identity: 57.202
alignment_block:
US-09-938-330-20 x BI081935
Align seg 1/1 to: BI081935 from: 1 to: 724

```

```

495 Alap[he]arg[thr]phe[glu]pro[cys]lys[glu]leu[trp]cys[ser]his[pro]as 511
|||||
1 GCGTTCGGGACCTTCGACCCCTGCACAAACAGCTCTGGTGCACACCAGCA 50
511 pasn[pro]tyr[phe]cys[lys]thr[lys]gly[pro]leu[asp]lythr[g] 528
|||||
51 CAACCCCTACTTTTGCAGACAAAGAGGGGTCTCCACTGATGATCTACTA 100
528 l[uc]yal[al]pro[glu]lystr[trp]cys[phel]ys[glu]his[cy]l[et]r[plu]ser 544
|||||
101 TGTGTGCGCCCTGCGACAGACGCTTCCAAAGGCCACTGATCTGAGTAC 150
545 pro[glu]in[thr]tyr[glu]ln[asp]gly[tyr]p[ser]ser[thr]lys[ph] 561
|||||
151 CCGCAT..ATCCTCAACGAGACGCACTGGGGTCTGCACTCCGTT 197
561 egl[ser]cys[ser]arg[ser]cys[glu]gly[yl]val[arg]ser[arg]ser[arg]s 578
|||||
198 TGGCTCTCTGCTACGACGACCTGTGGACAGCGCTGAAGTTCAAGACCGGCC 247
578 e[tyr]cys[asn]asn[pro]ser[pro]ala[tyr]glu[tyr]arg[pro]cys[leu]gly[pro] 594
|||||
248 AGTGTGACAAACCCACATCCAGCCAAATGGGGCGGCACTCTCGGGCCCTG 297
595 Meth[he]glu[tyr]glu[ln]val[cys]asn[ser]glu[glu]cys[pro]gly[thr]tyr[gl] 611
|||||
298 GCGATGACCTTCACGCTCTGCTCAACCCCAAGACTGTCCCAATTCCTTGC 347
611 uas[ph]e[arg]al[ag]ln[cys]al[al]ys[arg]asn[ser]tyr[tyr]val[his]g 628
|||||
348 TGACTTCGGGAGAGGAGCAGTGTCAAGAGCGAGCTGTACTTTGACCAAG 397
628 lna[asn]al[al]ys[his]ser[tyr]val[pro]tyr[glu]pro[asp]asp[ala]glu 644
|||||
398 GTGATGTCACAGCACCACTGCTGCTCATATGACACCGGATGCGCAAGAG 447
645 l[ys]cys[glu]leu[ile]cys[glu]ln[ser]ala[asp]th[glu]asp[yl]val[ph]eme 661
|||||
448 AGATGCCACCTCTACTGTGTGATGTCACAGAGACAGGGAGCTGTGCTAT 497
661 l[asn]glu[ln]val[his]asp[glu]tyr[arg]cys[ser]tyr[arg]asp[pro]tyr[s] 678
|||||
498 GAAGCGCATGTGTCACAGATGGAACAGCGTGTCTCTACAGACAGCGCTTCA 547
678 e[tyr]val[cys]ala[trg]glu[cys]val[pro]val[glu]cys[asp]lys[glu]val 694
|||||
548 GCGTGTGTGTGTGCTGGGACCTGACGAAAGTGCGATGTGATGCGCTGATC 597
695 gl[ys]er[met]lys[ala]asp[asp]lys[cys]glu[yl]val[cys]glu[yl]asp[asn]se 711
|||||
598 GGGTCCCGTAAAGCAGACAGCAAGTGTGGCGTGTGGGAGGCGACAAAC 647
711 r[his]cys[arg]thr[yl]al[lys]glu[tyr]thr[leu]glu[lys]al[ser]lys[glu]ala 727
|||||
648 CCACTGTCAACAGTGTGGAAGGCGACATTCACAGGTCAACCCAGCAAAACA 697
728 gl[yl]al[leu]lys[leu]val[glu]leu[pro] 736
698 GATTACATCAAGATGCTGTGAGATCCG 724
seq_name: gb_est1:BB635886

```

```

seq_documentation_block:
LOCUS BB635886 564 bp mRNA linear EST 31-AUG-2001
DEFINITION BB635886 RIKEN full-length enriched, 0 day neonate thymus Mus
musculus cDNA clone A430089F14 5', mRNA sequence.

```

file



```

ACCESSION      BE381517
VERSION        BE381517.1 GI:9326882
KEYWORDS
SOURCE         house mouse.
ORGANISM       Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus
REFERENCE      1 (bases 1 to 562)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cgabsr@email.nih.gov
                Tissue Procurement: Gilbert Smith, Ph.D.
                CDNA Library Preparation: Life Technologies, Inc.
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LNLN at:
                http://image.lnl.gov
                Plate: LHAM8787 row: h column: 05
                High quality sequence stop: 562.
FEATURES
    source
        location/Qualifiers
            1..562
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:3601588"
                /clone_lib="NCI CGAP Mam1"
                /tissue_type="tumor, biopsy sample"
                /dev_stage="10 months, virgin"
                /lab_host="DH10B"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
                Site:2: NotI; Cloned unidirectionally. Primer: oligo dT
                library constructed by life technologies. Investigator
                providing samples: Gilbert Smith, NIH"
BASE COUNT     106 a 179 c 156 g 121 t
ORIGIN
alignment_scores:
    Quality:   808.50          Length:   187
    Ratio:     4.870           Gaps:     2
    Percent Similarity:  88.770      Percent Identity: 76.471
alignment_block:
--09-938-330 x BE381517 ..
align seg 1/1 to: BE381517 from: 1 to: 562
324 GUAUgGlyAsnProSerArgSerLeuGlInValICysArGTrrAlahI 340
|||||
3 GAGATTGGAAACCCTCTCAGAGTGTCGAAGAATGTCTGCCGATGGCCATA 52
340 sserGlInArgLInAsPProSerHISalagIHSHISASPHISValY 357
||| |||||
53 CCTCGACAGAAAGCCGGACACAGCACGATGATGATCCACGAGCCATGCCA 102
||| |||||
357 alPhelEuThArGLInAsPphEgLYProSer.....GIYTyrALA 370
|||||
103 TCTTTCTCATCAGGCGAGATTGTGGCCCTCGGGGCATGCCAGGCTATGCT 152
|||||
371 ProValThrGLymETcySHSProlEuArGerCySaLaLeuAsnHisGI 387
|||||
153 CCGGTCACCGGGATGTCCACCCCTGCCGAGCTGTACTCAACCATGA 202
|||||
387 uASPolYPheSerSerAlaPheValIlleaIlaHISglUThrGLyHISValI 404
|||||
203 AGATGGCTTCCTCGGCATTGTGGTGCGTCAACGAGAGCGGCATGTCC 252
|||||
404 euGLyMeGLuHISASPolYGLnglYASncLYsaLaIsAsglUtHrSer 420
|||||
253 TGCGATGTGAGCATGACGAGGCAAGGCAACCGCTGCCGTGACAGAGTGCGG 302
|||||

```

```

421  Lengg1SerValMet1aProLeuVal1Gln1a1a1aPhe1iSarPhe1i 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
303  CTGGGACGACATCATGGCCCCCTGGGTACAGAGCGCTTCCTCATCGCTTCCA 352
437  sTTPSerArgCysSerLysLeuGluLeuSerAArgTyrLeuProSerTyrA 454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
353  CTGGTCCCGCTGCAGCCAGAGAGAGCTCAAGCCGCTACCTGCATTCCTTANG 402
454  spGys1LeuAaspAspProPheAspProAlaTyrPro_GlnProProG1 470
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
403  ACAGCGCTGGCGGATGATCCCTTGCCCCACGACGTGGCGACATTGGCTCA 452
470  uLeuProGly1LeaSTyrSerMetAspGluGlnCysArgPheAspPheg 487
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
453  ACCTCCAGGCGCTTCATATTCCATGACAGACAGATGCTGCTTGGACTTTGG 502
487  lySerGlyTyrGlnThrCysLeuAla1aPheArgThrPheGluProCysLys 503
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
503  GCCTGGGTTACATGATGTGCACGTGCTACCGACCTTCGACCCCTGCANA 552
504  GlnLeuTyr 506
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
553  CAGCTCTGG 561

seq_name: gb_est2:B1858020

seq_documentation_block:
LOCUS B1858020 767 bp mRNA linear EST 10-OCT-22
DEFINITION 6033848251 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:539351 5',
RNA sequence.
ACCESSION B1858020
VERSION B1858020.1 GI:15998767
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 767)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCMD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
plate: L1AM12003 row: a column: 08
High quality sequence stop: 764.
Location/Qualifiers
1..767
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:539351"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: breast; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC library."
BASE COUNT 239 a 158 c 177 g 193 t
ORIGIN

```

Percent Similarity: 75.000 Percent Identity: 56.818

alignment\_block:

US-09-938-330-20 x B1858020 ..

Align seg 1/1 to: B1858020 from: 1 to: 767

```

591 CysLeuGlyPromePhenGluTyrGlnValCysAsnSerGlnGluCysPr 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    4 TGCTGGGTAAATTGATGATGACGCTTTGATACACAGGAATGCGCA 53
607 oGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaTyrAsnSerT 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    54 AAAACCTTGAGACTTCAGACACAGCAGTGTGACGACGCAAACTCC 103
624 YTYrValHisGlnAsnAlaLysHisSerTyrValProTyrGluProAsp 640
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    104 ACTTGAATACCAATACCAACACCACTGGTGTGATGATGACATCCT 153
641 AspAspAlaGlnLysCysGluLeuLeuLeuGlnSerAlaAspThrGlyAs 657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    154 GACCCCAAGAAAGATGCCCTTACTGTCTAGTCCAGAGAGCTGAGAA 203
657 pValValPheMetAsnGlnValValHisAspGlyThrArgCysSerTyrA 674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    204 TGTGCTTACATGAAACAACTGTCATGATGAGAACGACGCTGCTTACA 253
674 rGAspProTyrSerValCysAlaArgGlyGluCysValProValGlyCys 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    254 AAGATCATATAGCATATGTGTGCGAGAGATGTGTGAAGTGGGCTGT 303
691 AspLysGlnValGlySerMetLysAlaAspLysCysGlyValCysGln 707
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    304 GATAAAGAAATGTGTTCTAATAAGGTGAGTAAGTGTGCTGTCTGTG 353
707 YGlyAspAsnSerHisCysArgThrValLysGlyThrLeuGlyLysAlaS 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    354 AGAGATTAATTCCTCCAGTCCGAAACCTGAAAGGAGCATTTACAGAACTC 403
724 eArgGlnAlaGlyAlaLeuLysLeuValGlnIleProAlaGlyAlaArg 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    404 CAGGAAGCTGGGTACCTTAAGATGTTGATATATACCCCTGGGGCTAGA 453
741 HisIleGlnIleGlnAlaLeuGlnLysSerProHisArgSerValVal 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    454 CATGCTTATATCCAGAGAGCGGCTCTCCTCATATCTTGCATATTA 503
757 sAsnGlnValThrGlySerPheIleLeuAsnProLysGlyLysGlnAlaT 774
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    504 GAACCGAGGCTACAGGCCATTATATTTAAATGGCAAGGGGAGGAGCA 553
774 hSerArgThrPheThrAlaMetGlyLeuGlnTropGlnLysAlaGln 790
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    554 AGTGGGAGCTTCAATGATCTGTGGTGTGAGGGGATATTAACATGAA 603
791 AspAlaLysGlnSerLeuLysThrSerGlyProLeuProGlnAlaIle.. 806
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    604 GATGACATTAAGATCTTCACACCGATGACCTTTATCATGATGACCTGACT 653
807 .....AlaIleLeuAlaLeuProProThrGlnG 816
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    654 GCTTGAATTAACCTCAAGAACATGATGACGCTCAAGCTGAGCA... 699
816 LysGlyProArgSerSerLeuAlaTyrLysTyrValIleHisGlnLysPleu 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    700 .....TATTAATGATCATCTCATCATTAAGAACTC 726
833 Leu..ProLeuIleGlySerAsnAsnValLeuLeuGlnGlu 845
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    727 TTGTACTTCAATCAACAGCAACAAATGTATCATCCAGAGAGAA 767

```

seq\_name: gb\_est2:Bf190891

seq documentation block:

LOCUS Bf190891 529 bp mRNA linear EST 02-NOV-2000

for late

DEFINITION 237553 MARC 2Pig Sus scrofa cDNA 5', mRNA sequence.

ACCESSION Bf190891

VERSION Bf190891.1 GI:11074260

KEYWORDS EST.

SOURCE pig.

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 529)

Fahrenkrug,S.C., Frenking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,

Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.,

and Keele,J.W.

Design and use of two pooled tissue normalized cDNA libraries for

EST discovery in swine

Unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smtche@mail.marc.usda.gov

Single pass sequencing. Bases called and alt-trimmed with phred

v0.980904.e. Vector identified by cross-match with the -minscore 18

and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGACG

Plate: 65 row: G column: 12

Seq primer: ATTGAGTGACATATAG.

Location/Qualifiers

1..529

/organism="Sus scrofa"

/db\_xref="taxon:9823"

/clone\_lib="MARC 2Pig"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;

Library made from pooled tissue from testis, ovary,

endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 103 a 163 c 158 g 105 t

ORIGIN

alignment\_scores:

Quality: 769.50 Length: 175

Ratio: 4.692 Gaps: 1

Percent Similarity: 93.714 Percent Identity: 78.286

alignment\_block:

US-09-938-330-20 x Bf190891 ..

Align seg 1/1 to: Bf190891 from: 1 to: 529

```

270 SerValValArgPheHisGlyLysGlnHisValGlnAsnTyrValLeuTh 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    3 TCTGTGGTCCAGTCCAGGAGAGGACAGCAGTGTGAGAACTGCTGAC 52
286 rLeuMetAsnIleValAspGluIleTyrHisAspGlnSerLeuGlyValH 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    53 TCTCATGACATCGTCAATGATGATTTATCATGAGACTCCCTGGGGCCC 102
303 isIleAsnIleAlaLeuValArgLeuLeuMetValGlyTyrArgGlnSer 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    103 ACATCAATGTGCTGTGCGGATTAATCTGCTGAGATTCAGGGAATCC 152
320 LeuSerLeuIleGluArgGlyAsnProSerArgSerLeuGlnGluValCy 336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    153 ATGAGCCTCATTTGATCGGAAACCTCTGACAGTGTGAGAACTGCTG 202
336 sArgTyrPalaHisSerGlnGlnArgGlnAspProSerHisAlaGlnHis 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

203 CCGCTGGGCTTACCTGCACAGACGCGGACACTGGCCACGAGAGTAC 252  
 353 iAspHisValValPheLeuThrA:ggInaspPheGlyProSer..... 367  
 253 AGGATCATGTCATTTCTCTCAACCGCGAGACTTTGGCCCTCGGGCATG 302  
 368 ...GlyTYrAlaProValThrGlyMetCysHisProLeuArgSerCysAl 383  
 303 CAAGCTATGCTCTGTCACCGGGATGTCACACCTGTCGCCACTGCAC 352  
 383 aleuasnHisgInuAspGlyPheSerSerAlaPheValIleAlaHisGluT 400  
 353 CTGGAACCAACGAGATGGCTTCTCTCAGCGTGTGTCGGCCATGAGA 402  
 400 hGlyHisValLeuGlyMetGlnHisAspGlyGlnGlyAsnGlyCysAla 416  
 403 CTGGCCATGTGCTGGGCGATGAGCATGACGGGCGAGCGCACCGTGGCGC 452  
 417 AspGluThrSerLeuGlySerValMetAlaProLeuValGlnAlaAlaPhe 433  
 453 GACGAGGTCGGCGTGGCGAGCATGACACCGCTGCTGCGAGGACGCTT 502  
 433 eHisArpHeHisTrpSerArgCys 441  
 503 CCACCGCTTCCACTGCTGCCCTGC 527

seq\_name: gb\_estl:AJ403134

seq\_documentation\_block:

LOCUS AJ403134 418 bp mRNA linear *day late*  
 DEFINITION AJ403134 3.3 (upregulated in larynx carcinoma) Homo sapiens cDNA

ACCESSION AJ403134

VERSION AJ403134.1 GI:9903409

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 418)  
 Frohme, M., Scharm, B., Delius, H., Knecht, R. and Heideisel, J.D.  
 Use of representative difference analysis and cDNA arrays for  
 transcriptional profiling of tumor tissue

TITLE

JOURNAL Ann. N. Y. Acad. Sci. 910, 85-104 (2000)

MEDLINE 20369561

COMMENT Contact: Frohme M

FUNCTIONAL GENOME ANALYSIS  
 Deutsches Krebsforschungszentrum  
 INF 506, 69120 Heidelberg, Germany

EMAIL: m.frohme@dkfz-heidelberg.de.

FEATURES

source

1. 418  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="12"

/clone\_lib="3.3 (upregulated in larynx carcinoma)"  
 /tissue\_type="larynx carcinoma"

BASE COUNT 71 a 149 c 120 g 78 t

ORIGIN

alignment\_scores:

Quality: 767.00 Length: 139  
 Ratio: 5.518 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.281

alignment\_block:  
 US-09-938-330-20 x AJ403134 ..

Align seg 1/1 to: AJ403134 from: 1 to: 418

323 TiegLuArgGlyAsnProSerArgSerLeuGlnValCysArgTrpAl 339

|||||

2 ATCGAGCGGGAGACCCCTCACGACGCTGAGAGGTGTGTGCTGGGC 51  
 339 aHisSerGlnGlnArgGlnAspProSerHisAlaGlnHisHisAspHis 356  
 52 ACATCTCCAGACGCGCCAGAGACCCAGCAGCAGCTGAGACCATACACAG 101  
 356 aIValPheLeuThrArgGlnAspPheGlyProSerGlyTYrAlaProVal 372  
 102 TTGTGTCTCTCACCCGCGGACCTTGGCCCTCAGGATGATGACCCGTC 151  
 373 ThrGlyMetCysHisProLeuArgSerCysAlaLeuasnHisGlnuAspG 389  
 152 ACTGCAATGTGTACACCCCTCAGAGAGCTGTGCTCAACATGAGATGG 201  
 389 yPheSerSerAlaPheValIleAlaHisGluThrGlyHisValLeuGly 406  
 202 CTTCCTCTCACGCTTGTGATAGCTCAGAGACGCGGCGAGCTGTGGCA 251  
 406 eGlnHisAspGlyGlnGlyAsnGlyCysAlaAspGluThrSerLeuGly 422  
 252 TGAACATGACGTCAGGCGGATGCTGTGACATGAGACACGCTGGGC 301  
 423 SerValMetAlaProLeuValGlnAlaAlaPheHisArpHeHisTrpSe 439  
 302 AGGTCATGAGCGGCCCTGTGTGACAGCTCTTCCACCGCTTCATTGGTC 351  
 439 rArgCysSerIysLeuGlnLeuSerArgTYrLeuProSerTYrAspCysAl 456  
 352 CCGCTGACGACGAGCTGAGGAGCTCAGCGCTACCTCCCTCAGACTGGC 401  
 456 aleuAspAspProPhe 461  
 402 TCTCTGATGACCCCTTT 418

seq\_name: gb\_est2:BJ096904

seq\_documentation\_block:

LOCUS BJ096904 555 bp mRNA linear *Feb 16 to*  
 DEFINITION BJ096904 NIBB Mochi normalized Xenopus early gastrula library

ACCESSION BJ096904

VERSION BJ096904.1 GI:17598016

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM Xenopus laevis

REFERENCE 1 (bases 1 to 555)  
 Kitayama, A., Terasaka, C., Mochi, M., Ueno, N., Shin-i, T. and Kohara, Y.

EXPRESSED GENES IN X. LAEVIS EMBRYO  
 Unpublished (2001)  
 Contact: Tadasi Shin-i  
 Center for Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp.

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 555  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="XL15118"

/clone\_lib="NIBB Mochi normalized Xenopus early gastrula library"  
 /tissue\_type="whole embryo"  
 /dev\_stage="stage 10.5"

BASE COUNT 159 a 120 c 131 g 145 t

ORIGIN

## alignment\_scores:

Quality:	748.00	Length:	142
Ratio:	5.582	Gaps:	0
Percent Similarity:	94.366	Percent Identity:	84.507

## alignment\_block:

US-09-938-330-20 x BJ096904 ..

Align seg 1/1 to: BJ096904 from: 1 to: 555

```
452 SerTyrAspCysLeuLeuAspProPheAspProAlaTrpProGlnPr 468
|||||
129 TCATATGACTGCTTCTGTGATGACCCCTTCGATCACAAATGGCCCAATT 178
|||||
468 oProGluLeuProGlyIleAsnTyrSerMetAspGluGlnCysArgPhea 485
|||||
179 ACCAGAACTCCAGAAATCATATTCATGGATGACAAATGAGATTGCG 228
|||||
485 sPheGlySerGlyTyrGlnThrCysLeuAlaPheArgThrPheGluPro 501
|||||
229 ATTTCGTTGGGGCTACAGATGTCACCTCTTCAGACTTTTGACCT 278
|||||
502 CysLysGlnLeuTrpCysSerHisProAspAsnProTyrPheCysLysTh 518
|||||
279 TGCAGCGAGTTGTGTGACAGCCAGATATCTTATTTCGCAAAAC 328
|||||
518 rLysLysGlyProProLeuAspGlyThrGluCysAlaProGlyLysTrpC 535
|||||
329 GAAGAAAGGCCCTCCACTGTATGAGACAGAAATGCTCCCGAAAGTGGT 378
|||||
535 yspheLysGlyHisCysIleTrpLysSerProGluGlnThrTyrGlyGln 551
|||||
379 GCTTTAAAGGATATTCATTGGAAAGACACCAAGAGAGTCTTACAGTCAG 428
|||||
552 AspGlyLysTrpSerSerTrpThrLysPheGlySerCysSerArgSerCy 568
|||||
429 GATGAGAGCTGAGACTCATGTGTCAGAGTTGATCTTCTCGAACCTG 478
|||||
568 sGlyGlyGlyValArgSerArgSerArgSerCysAsnAsnProSerProA 585
|||||
479 TGGAGGTGAGAGTGCCTCCCGTAGCAGACAAATGCAATMACCAGAGCAT 528
|||||
585 laTyrGlyGlyArgProCysLeuGly 593
:::|||||
529 CATATGGAGAGCGGATTTGCCCTGCT 554
```





```

137 ...GlnGlnAspPheArgGlnPheArgGlnProLeuArgGlnLucy 152
1222 .....
427 CAGCCCGATGCACTGGGTGGTCACTGAGGACACACTTTGGAGACTG 476
152 sValTyrThrGlyGlyValThrGlyMetProGlyAlaAlaAlaAlaIles 169
477 CTGCTACACAGGAGAGAGTGGGGATATGCAAGCTCTGGGTGTCATCT 526
169 eAsnGlyAspGlyLeuAlaGlyLeu...IleArgThrAspSerThrAsp 184
527 GCACCTCTCTGGGTGCTGAGAGCTTGGTGTCTGACCCAGAGAGAGAGC 576
185 PheHeliGlu.....ProLeuGluArg 192
577 TATACCTGGAGAGAGGGGCTGGGGACCTTCAAGGCTCTCCATTAATTC 626
192 gGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 209
627 GCGAATCCAAAGATCTCCAGCTGCGAGGCGACACCTGTGCTGAGCTGCG 676
209 rGlnGlnAlaVal.....GlnGlnGlnGlnGlnGlnGlnGlnGlnGln 223
677 GGGATCTGTACACATCTCAGACGCGACACAGACGACCCCTGGGACAGCG 726
224 HisAsnGlnAlaPheGlyLeuGlyAspLeuProAsnLeuGlnGlyLeu 240
727 CAC..... 729
240 lGlyAspGlnLeuGlyAspThrGlnArgGlyArgGlyHisAlaLysProG 257
730 .....ATTGCGCGAGGCGGGATGTGTAACAG 757
257 lYSerTyrSerIleGlnValLeuLeuValValAspAspSer...ValVal 272
758 AGACCAAACTGTGAGTGGTGGTGTGCTGATCTCGAGAGGCCAG 807
273 ArgPheHisGlyLysGlnHisValGlnAsnGlyValLeuThrLeuMetAs 289
808 AAATACCGGAGCTTCCAGACCTGCTAAACCGACACTGGAAGTGGCCCT 857
289 nIleValAspGluIleTyrHisAspGluSerLeuGlyValHisIleAsnI 306
858 CTTCGCTGACATCTTCCGCGCC.....CTGAATGTACAG 895
306 lAlaLeuValArgLeuIleMetValGlyTyrArgGlnSerLeuSerLeu 322
896 TGGCACTAGTGGCTGGAGGCTGACCCAGCAGCTGACCTGGTGAATC 945
323 lIleGluArgGlyAsnProSerArgSerLeuGlnGlnValCysArgTyr.. 338
946 .....AGCCCAAAACCCAGCTGACCTCGAATACTCTCCACTGCGG 989
339 ...AlaHisSerGlnGlnArgGlnAspProSerHisAlaGlnHisIleSA 354
990 CAGGGCACTTGGCTGCTGCTGATGCC.....CATTG 1021
354 sPheIleValAlaPheLeuThrArgGlnAspPhe...GlyProSer...Gly 368
1022 ACAGTCCCGCAGCTGTGACTGTACTTCTCTGGGCTTCCAGGTGGGC 1071
369 TyrAlaProValThrGlyMetCysHisProLeuArgSerCysAlaLeuAs 385
1072 ATGGCAATTCAGAACTCATCTGTCTCTGACTTCTCAGAGAGTGTGAA 1121
385 nHisGluAspGlyPheSer.....SerAlaPheValIleAlaHisG 399
1122 CATGCAACCACTCCACGATCTCCGGAGTCCGCTTCCATTAAGCCCATG 1171
399 luthrGlyHisValLeuGlyMetGlnHisAspGlyGlnGlnLysGlnGly 415
1172 AGTGGGCGCACAGCTGGGCTGGACCATGATTTGCTGGGAATAGCTGC 1221

```

```

416 AlaAspGlnThrSerLeuGlySerValMetAlaProLeuValGlnAlaI 432
1222 .....CCCTGTCCAGGTCCAGC 1238
432 aPheHisArgPheHisIleThrSerArgCysSerLysLeuGlnLeuSerArg 449
1239 C..... 1239
449 YrLeuProSerTyrAspCysLeuLeuAspAspProPheAspProAlaTyr 465
1240 .....CCAGCCAAAGACTGCATCATGAGAGGCCCTCCACAGAC..... 1275
466 ProGlnProProGlnLeuProGlyIleAsnTyrSer..... 477
1276 .....TTCTTACAGAGCTTGAACCTTACAGCACTGACGAGCGAGC 1313
478 .MetAspGlnGlnCysArgPheAspPheGlySerGlyTyrGlnThrCysL 494
1314 GGCCCTGAGAAAGCCCTCTGAT.....GGAATGGGCACTGCC 1354
494 euAlaPheArgThr..... 498
1355 TCTTGCAGAGGCTGCTACCTACCCCTATGCTGCTTTCGGGAAAT 1404
499 .....PheGluProCysLysGlnLeuThrCysSerHisProAspAsnPr 513
1405 ATGTTTGTGAGAGCGGGCGAGCATGTGACTGTGCTTCTGAGATGAC.. 1452
513 oTyrPheCysLysThrLysGlyGlyProProLeuAspGlyThrGlnGly.. 529
1453 .....TCC.....GTGATCCCTGCTGTGATTTCTTGACTGCTGC 1486
530 ....AlaProGlyLysTyrCysPheLys...GlyHisCysIleTyrLys 543
1487 AGGTGAGGCGAGGTGCACATGTGCATCTGACGAGCCCTGTGT..... 1530
544 SerProGlnGlnThrTyrGlnGlnAspGlyGlyIleThrSerSerThrLys 560
1531 .....CAAATTCGACAGCTGCGCCGCTGCTGCTGCTGCTGCTGCTGCT 1574
560 sPheGlySerCysSer.....ArgSerCysGlyGlyAlaValArgSera 575
1575 CAGAGGAGATGTGACTTGGCTGAATTCGCCACAGAGAC..... 1614
575 rGSerArgSerCysAsnAsnProSerProAlaTyrGly...GlyArgPro 590
1615 ...AGCTTCCAGTGT...CCCCGTGATGTCAAGCTTACGGGATGCGAGGCC 1659
591 CysLeuGlyProMetPheGluTyrGlnValCysAsnSerGlnGlnCysPr 607
1660 TGGCTGTGGCGGCA.....GCTGTGTGATGCACAGCGGCTGTGT.. 1698
607 oGlyThrTyrGlnAspPheArgAlaGlnGlnCysAlaLysArgAsnSer 624
1699 .GCCTCTAT.....GCCAGCAGTGC..... 1719
624 yTyrValHisGlnAsnAlaLysHisSerTyrValProTyrGluProAsp 640
1720 .....CAGTCACTTGGGAGCT..... 1737
641 AspAspAlaGlnCysGlnLeuIleCys...GlnSerAlaAspThrGln 656
1738 ...GGAGCCAGCGCGCTGGCGCACCTTGGCTCCAGACACTAATCTCG 1784
656 yAspValValPheMetAsnGlnValValHisAspGlyThrArgCysSer 673
1785 GGAAATGCTTT..... 1797
673 yTyrArgAspProTyrSerValCysAlaArgGlyGlnCysValProValGly 689
1798 .....GGGACTGT.....GGG 1809
690 CysAspLysGlnValGlySerMetLysAlaAspAspLysCysGlyValCys 706

```

```

1810 CGCAACCCGAGTGTATGTCTCTGACCCCTAGAGATGCATTTG 1859
706 sgllylaspasnserhisacyarthrvalylsrglythrleuglylva 723
1860 TGGG.....CAGCTCCAGTGCCAGACA.....GGTAGGA 1888
723 laserlysglnalaglalaaleuylsleuvalglnlleproalaglala 739
1889 CCCAGCCTCTGCTGGCTCCATC..... 1911
740 ArgHisIleGlInleGlAlaLeuGlulysSerProHisArgSerValva 756
1912 CGGAGATCTACTCTGGAGACAATAGAT..... 1938
756 llysaasnlnvalthrlyserphenlleuasnprolysglylsglua 773
1939 .....GTGATGCG..... 1947
773 lathrserargthrphenrhalametylleuglulrpglu.....asp 787
1948 .....ACTGAGCTGAACTGCAGCTGGGTGCACCTGAC 1980
788 Alavalgluasplalalysgluserleuylsrserglyproleuprog1 804
1981 CTGGCAGATGATGTGGCCCGCCCTC..... 2007
804 ualalalealaleualeuaprothrprothrleuglylproargser 821
2008 .....CTGACTCTCTCCGCGACAGACCTGTGGCCCT..... 2037
821 erleualatylrlystyvallehisglu...aspleuileuproleuile 836
2038 .....GGCCTGTGTGTATAGACATCATGATCCAGCCGTGATCTCTCTG 2082
837 glyserasnasnvalleuenglunetaspthrlyrgrlulrpalale 853
2083 GGGCGCACAGAA..... 2094
853 ulyssertrpalaprocysserlysalacys...glyglylileglinp 869
2095 .....TGTGAGAGCAATGCCATGAGCATGGGTCTCTGTG 2128
869 heThrlystyryllys ArgArgArgArgasphishismetvalglinar 885
2129 ACAGCAAGAGCAGCTGCTACTGTAGAGGGCTGGCAGACCCCTGACTGC 2178
885 gHisleucysasphishlylsyArgrProlyProIleargArgArgcysa 902
2179 ACCACTGAGCTCAAGACACAGCTCCCTGACA..... 2212
902 snGlnHisProcysserGlnProValThrValThrGlnulrpglylala 918
2213 .....CAGGGCTGCTCCACAGCTCTGCTGTATGTGCTCTG..... 2251
919 CysSerArgSerCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCy 935
2252 .....TGATG 2256
935 sLeuLeuProleuSerasnGlyThrHisLysValMetProAlaLysalac 952
2257 CTGTGTGTGCA...CTACTGTGTAC.....GTCCCGGCT 2288
952 ysAlaGlyAspArgProGlnAlaArgArgProCysLeuArgValProCys 968
2289 GCACGAGGAGCTTGCCAGCTCAAGGAGCA..... 2320
969 ProAlaGlnTrpArgLeuGlyAlaTrpSerGlnCysSerAlaThrCysgl 985
2321 CCTGCC.....AGTACAGGGCAGACCCCAATCTGCTCTCT 2355
985 yGlnGlyIleGlnGlnArgGlnValValCysArgThrAsnAlaAsnSerL 1002

```

```

2356 GAAAGCGCAGACCTCCGCAAGAGGCTCTGTGCCAGCAGCATAAAGC 2405
1002 euGlyHisCysGlnGlyAspArgProAspThrValGlnValCysSerleu 1018
2406 TA...GTGCTCTCAGCTTCCGAGCCCTTCCAGAGCCCGCGGCTG 2452
1019 ProAlaCysGlyLysAsnHisGlnAsnSerThrValArgAlaAspValTr 1035
2453 ACCCTGTGT.....CCAAAGACTCCAGTCTCAG..... 2482
1035 pGluLeuGlyThrProGlnGlyInlTrpValProGlnSerGlyProleu 1052
2483 .....GCCCAAGCAAGCC.....C 2498
1052 lAsProIleasnlyslleSerSerMetCysAlaAlaGluProCysThrly 1068
2499 ACCCC.....CAAGGAGCCACTGCTGCTCCAGACCCCAAGGCGCGT 2539
1069 AspArgSerValPheCysGlnMetGluValLeuAspArgTrpCysSeril 1085
2540 GCCCATCGGTGACCTGCCCGCCAGGAGGCTGAA.....TCCCG 2580
1085 eProGlyTyf.....HisArgLeuCysCysValSerc 1096
2581 CCCCTAGTGTACCTCCAGACAGCCAGCCAGCTC.....CGACAGT 2624
1096 ysIleLysLysAlaSerGlyProAsnProGlyProAspProGlyProthr 1112
2625 GTCCTGCTCTACTCTGACTCTCGGAGGT..... 2656
1113 SerLeuProProPheSerThrProGlySerProleuProGlyProGlnas 1129
2657 .....TCCGCTGCCCTCAAGCGGACTTAAAGGCTTCA.. 2689
1129 pProAlaaspAlaalaGluProGlyLysProThrGlySerGlnasp 1145
2690 .....GAGCGGGCGTCCCTGTGAGTCCCTACCATGACGAAAGC 2732
1146 HisGlnHisGlyArgAlaThrGlnLeuProGlyAlaLeuAspThrSerc 1162
2733 GCCAAGACTGCGGTGTAAAGACTCCGGC.....ACCGCAC 2773
1162 rProGlyThrGlnHisProPheAlaProGluThrProIleProGlyAlas 1179
2774 GCGCTGTAAAGCAACACTCTCGGACCTGCCCGGCTGTGACAGCGGGG 2823
1179 eTrpSerIleSerProThrThrProGlyLysLeuProTrpGlyTrpThr 1195
2824 CTGG.....GAGGCGCTGGGGGTGGAGC 2849

seq_name: /cgn2_6/ptodata/2/lna/6B_COMP.seq:US-09-920-048-1
seq_documentation_block:
? Sequence 1, Application US/09920048
? Patent No. 6344352
? GENERAL INFORMATION:
? APPLICANT: MERKULOV, Gennady et al
? TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND
? TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
? FILE REFERENCE: CL001177DIY
? CURRENT APPLICATION NUMBER: US/09/920,048
? PRIOR APPLICATION NUMBER: 09/813,819
? PRIOR FILING DATE: 2001-03-22
? NUMBER OF SEQ ID NOS: 4
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 1
? LENGTH: 2968
? TYPE: DNA
? ORGANISM: Human
US-09-920-048-1

```

## alignment\_scores:

Quality: 310.00 Length: 1268  
 Ratio: 0.598 Gaps: 73  
 Percent Similarity: 40.852 Percent Identity: 22.082

## alignment\_block:

US-09-938-330-20 x US-09-920-048-1 ..

Align seg 1/1 to: US-09-920-048-1 from: 1 to: 2968

```

3  ProLeuArgAlaLeuSerTyrLeuLeuProLeuHisCysAlaLeu.. 18
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55  CCGGCCGGAGCGCGCGCTGCATGCGGCTGCGCTGCGGCGCTGGC 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19  CysThrAlaAlaGlySerArgThrProGluLeuHisLeuSerGlyLysL 35
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 GCTCTGGGCGCGCGCGCGCTGCGCTGCGCTGCGCGCTGCCAAATATAG 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35  euSerAspTyrGlyValThrValProCysSerThrAspPheArgGlyArg 51
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 CCTTG.....CTGCGATGCC.....TCA 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
52  PheLeuSerHisValSerGlyProAlaAlaSerAlaGlySerMe 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 GTACTGTCTGGGGTCTCTGGGACCTGCAGTGGC..... 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68  lValValAspThrProThrLeuProArgHisSerSerHisLeuArgV 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
211 .....ACTGAGAGCAGCAGCAGCAGCAGCAGCAGAGAG 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
85  aAlaArgSerProLeuHisProGly..... 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
242 CCCGAGGAGCGCTTGGAGCCCGCAGCTCTTCAAGAGCATCTCCCAT 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94  .....GlyThrLeuThrProGlyArgValGlyArgHis 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
292 ACCCTCAAAAGCTGCTCAGACAGCTGCTGCT.....GA 326
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104  sSerLeuTyrPheAsnValThrValPheGlyLysGluLeuHisLeuArgL 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
327 GCCCTGAGAGTCAAGTGGAGCTGAGCTGAGCTGATATCTCTGGAGC 376
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121  euArgProAsnArgArgLeuValValProGlySerSerValGluTrp... 136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
377 TGTCTACAGATATAGGAGTGTGTCCAGCGCGCCAAACCTGTGTGTCAG 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
137  ...GlnGluAspPheArgGluLeuPheArgGlnProLeuArgGlnGly 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
427 CAGCCGATGGCAGCTCGGGTGGTCACTGAGGAGCAGACACTTTGGAGACTG 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
152  sValTyrThrGlyGlyValThrGlyMetProGlyValAlaAlaValAla 169
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
477 CTGCTACACGAGGAGAGTGGGAGATGCAAGGCTCTCGGTGTCATCT 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169  eArgAsnCysAspGlyLeuAlaGlyLeu...LeaArgThrAspSerThrAsp 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 GCACCTGCTCTGGCTCAGAGGCTTGGTCTCTGACCCAGAGAGAAC 576
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185  PhePheIleGlu.....ProLeuGluArg 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
577 TATACCTTGGAGCAGCGCGCTGGGACCTTCAAGGCTCTCCATTAATTTTC 626
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
192  gGlyGlnGlnGlnLysGlnAlaSerGlyArgThrHisValValTyrArgA 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
627 GCGAATCCAGAGATCTCCACCTGCGAGGCCACACTGTGCTGAGCTGGC 676
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
209  rGlnLysAlaVal.....GlnGlnGlnLysLysLysLysLysLysLys 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
677 GGAATATCTGTACACTCAGAGCCACAGAGACCCCTGGGACAGCGCC 726
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224  HisAsnGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

727  CAC..... 729
240  lGlyAspGlnLeuGlyAspThrGlnArgLysArgHisAlaLysProG 257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
730  .....ATTGCGGAGGCGGGGATGCGTAAACAG 757
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
257  LysTyrSerIleGlnValLeuValValAspSer...ValVal 272
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
758  AGACCAAGACTGTGATGTGTGATGTGTGATGTGTGATGTGTGATGTG 807
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
273  ArgPheHisGlyGlnHisValGlnAsnTyrValIleThrLeuMetAs 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
808  AAATACCGGAGCTTCCAGCAGCTGTAAACCGCACAGGAGTGGCCCT 857
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
289  nIleValAspGluIleTyrHisAspGluSerLeuGlyValHisIleAsn 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
858  CTGTGCGACACTTCTTCCGCC.....CTGAATGTACAG 895
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
306  lValLeuValArgLeuIleMetValGlyTyrArgGlnSerLeuSerLeu 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
896  TGGCACTAGTGGGCTGAGGCTGAGCCAGCCAGCTGAGCTGAGATC 945
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
323  lIleGlnArgGlyAsnProSerArgSerLeuGlnValValCysArgTrp.. 338
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
946  ....AGCCCAAAACCCAGCTGTACCTCGAAACTTCTTCCACTGGCG 989
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339  ....AlaHisSerGlnLysArgGlnAspProSerHisAlaGlnHisHis 354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
990  CAGGCAACATTTGCTGCTCGATTTGCC.....CARG 1021
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354  sPheValValPheLeuThrArgGlnAspPhe...GlyProSer...Gly 368
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1022  ACAGTGCCAGCTGTGACTGTACTTCACTTCTCTGCGCTAGCGTGGCC 1071
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369  TyrAlaProValThrGlyMetCysHisProLeuArgSerCysAlaLeuAs 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1072  ATGGCCCTTGAAGATCATCTTCTTCCGCACTTCTCAGAGGTGTGA 1121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
385  nHisGluAspGlyPheSer.....SerAlaPheValIleAlaHisG 399
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1122  CATGACCACTCCACAGCATCTCTGGAGTGGCTCTCTCCATAGCCCATG 1171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399  lThrThrGlnHisValLeuGlnMetGlnHisAspGlyGlnGlnLysLys 415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1172  AGTTGGCCACAGCTGTGGCTGACCATATTTGCTGTGGAAATACCTGC 1221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416  AlaAspGluThrSerLeuGlySerValMetAlaProLeuValGlnAlaAl 432
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1222  .....CCCTGTCCAGGTCCAGC 1238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432  aPheHisArgPheHisThrSerArgCysSerLysLeuGlnLeuSerArgT 449
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1239  C..... 1239
449  yIleuProSerTyrAspCysLeuLeuAspAspProPheAspProAlaTrp 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1240  ....CCAGCCAAAGACTGCATCTGAGAGGCTCCACAGAC..... 1275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
466  ProGlnProProGluLeuProGlyIleAsnTyrSer..... 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1276  .....TTCTACACAGGCTGAGCTTCACTGCAACATGCAAGCCGAGC 1313
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
478  MetAspGlnGlnCysArgPheAspPheGlySerGlyTyrGlnThrCysL 494
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1314  GGCCTGTGAGAAAGCCCTCTGTGAT.....GGAATGGCAGCTGCC 1354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
494  euAlaPheArgThr..... 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1355  TCTTTCGAACGAGCTGCTTACCTACCCCTAAGCTGCTTCTTGGGGAAT 1404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
499  ....PheGluProCysLysGlnLeuTrpCysSerHisProAspAsnPr 513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1405  ATGTTTGTGAGCCGGGAGCAGTGTGATGTGTGCTTCTGATGAC.. 1452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

513 oTyrrhPheGySLystrHrLylsYSLpPrPoleuAspGlyThrGluCys. 529  
1453 .....TGC.....GTCAATCCGTCGCTGTGATCTTTGACCTGCC 1468  
530 .....AlaProGlyLysTrpCysPheLys..GlyHisCysIleTrpLys 543  
1487 AGCTGAGGCCACAGGTGCACAGCTGCATGCATGACGAGCACCCTGTT. 1530  
544 SerProGluGlnThrTyrgLysGlnAspGlyGlyTrpSerSerTrpThrLy 560  
1531 .....CAAAATTGCCAGTCGCGCCGCTGGCTGGCTGCACATGTCGTCAC 1574  
560 sPheGlySerCysSer.....ArgSerCysGlyGlyValArgSerA 575  
575 CAGAGGGGATTTGTACTTGCCTAAATTGCTCCAGAGAC..... 1614  
575 rGSerArgSerCysAsnAsnProSerProAlaTrpLy..GlyArgPro 590  
1615 ..ACCTCCACAGT...CCCCGTGATGTCAGAGCTAGGGATGGCAGGCC 1655  
591 CysLeuGlyProMetPheGlyTrpGlnValCysAsnSerGluCysPr 607  
1660 TGCCCTGCGGGCAA.....GCTGTGTGATGACGACGGCGCTTT.. 1698  
607 oGlyThrTyrgLysAspPheArgLagLysGlnCysAlaLysArgAsnSer 624  
1699 .GCCTCTAT.....GCCAGCAGTGC..... 1719  
624 yTrpValHisGlnAsnAlaLysHisSerTrpValProTyrgLysProAsp 640  
1720 .....:::..:::..:::..:::..:::..:::..:::..:::..:::.. 1737  
641 AspAspAlaGlnCysGluLeuIleCys..GlnSerAlaAspThrGlu 656  
1738 ..GGAGCCAGCCCGCTGCGCACCTTGTCCCGACAGCTAATACTCG 1784  
656 yAspValValPheMetAsnGlnValValHisAspLysThrArgCysSer 673  
1785 GGGAAATGCTTT..... 1797  
673 yTrpAspProTyrgSerValCysAlaArgGlyGluCysValProValGly 689  
1798 .....GGAGCTGT.....GGG 1809  
690 CysAspLysGluValGlySerMetLysAlaAspAspLysCysGlyValCy 706  
810 CGCAACCCAGTGGCGAGTATGTGCTGCCACCCCTAGAGATGCCATTG 1859  
706 sGlyGlyAspAsnSerHisCysArgTrpThrValLysGlyThrLeuGlyLysA 723  
1860 TGGG.....CAAGTCAGATGGCAGACA.....GGTGA 1888  
723 lAserLysGlnAlaGlyLysLeuLysLeuValGlnIleProAlaGlyAla 739  
1889 CCGAGCTCTCTGGGCTGCATC..... 1911  
740 ArgHisIleGlnIleGlnAlaLeuGluLysSerProHisArgSerValVa 756  
1912 CGGGATCTACTCTGGGAACAATAGAT..... 1938  
756 lLysAsnGlnValTrpGlySerPheIleLeuAsnProLysGlyLysGlnA 773  
1939 .....GTGANTGG..... 1947  
773 lathSerArgThrPheThrAlaMetGlyLeuGlnTrpGlu.....Asp 787  
1948 .....ACTGAGCTGAACCTGACAGCTGGGTGCACCTGAGAC 1980  
788 AlaValGluAspAlaLysGlySerLeuLysThrSerCylProLeuProGlu 804  
1981 CTGGGAGTGAATGTGTGCCACCCCTC..... 2007

[illegible]

```

1113 SerLeuProProHeserThrProGlySerProLeuProGlyProGlnAs 1129
2657 .....TCCGCTGCTCCAGCGCGACTTAAAGGCTTCA.. 2689
1129 ProAlaAspAlaAlaGluProProGly LysProThrGlySerGluAsp 1145
2690 .....GAGCGGGCGGCTGCTGAGACTCCCTTACATGACTGAAAGC 2732
1146 HisGlnHisGlyAlaGluAlaThrGlnLeuProGlyAlaLeuAspThrSer 1162
2733 GCGAGAGACTGGCGGTCTTAAAGACTCGGCG.....ACCGCCAC 2773
1162 rProGlyThrGlnHisProHeserProGluThrProLeuProGlyAla 1179
2774 GCGCTGTCAAGCAACACTGCGGAGCTGCGGCGGTGTCAGCGCGG 2823
1179 eTrpSerIleSerProThrThrProGlyGlyLeuProThrGlyThr 1195
824 CTGG.....GAGGGGCTGGGGGTGGACG 2849
seq_name: /cgn2/6/ptodata/2/ina/6A_COMB.seq: US-08-985-526-4

seq_documentation_block:
; Sequence 4, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; NUMBER OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Boyle, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1326 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-985-526-4

alignment_scores:
Quality: 297.00 Length: 587
Ratio: 1.356 Gaps: 24
Percent Similarity: 37.308 Percent Identity: 20.443

alignment_block:

```

```

US-09-938-330-20 x US-08-985-526-4
Align seg 1/1 to: US-08-985-526-4 from: 1 to: 1326
473 GlyIleAsnTyrSerMetAspGluGlnCysArgPheAspPheIleArgI 489
64 GAGTGTACATACAGAAATTAACAGAGAAATGAGATGTGTAT.....AG 104
489 yTyrGlnThrCysLeuAlaPheArgThrPheGluProCysGlyGlnLeu 506
105 CTGCACTGAGTGTACTGTCTGCAAACTAGTTACCATGTGCAAAAGGTCT 154
506 rPcysSerHisProAspAsnProTyrPheCysLysThrLysGlyPro 522
155 CCTGC.....CCCATATGCC.....TGCTTCATATGCGACAGTTCT 192
523 ProLeuAspGlyThrGlnCysAlaProGlyLysThr..... 534
193 .....GATGA...GAATGCTGTCTGCTGTGTGGCCAGGCACTGTGC 233
535 ..... 535
234 GAGAGATGCTGCTGCTCCATGCTCCAGTGCAGTGCCTCTGTTCTACAGCT 283
535 yspheLysGly..... 538
284 GTGGCAATGCAATTTCAGCAGCGCGCGCTCTGCAATAGCTTCACAAAC 333
539 .....HisCysIleThrLysSerPr 545
334 CGATGTGAGGGCTCTCGTCCAGACAGCGACTGACCATTCAGAGAGTG 383
545 oGluInThrTyrGlnAspGlyIleTyrSerSerThrLysPheG 562
384 TGACAAAGATTTAACAGAGATGCGGTGAGCGACAGTGCCTGCGGTGT 433
562 LysCysSerArgSerCysGlyGlyValArgSerArgSerArgSer 578
434 CATCTTGTCTGTGACATGTGTGATGTGATGATCAGAGATCGGCTC 483
579 CysAsnAsnProSerProAlaTyrGlyAlaArgProCysLeuGlyPro 595
484 TCCAACTCTCCAGCCGCCAGATGATGGAACCTTGAAGGGAAGC 533
595 rPheGluTyrGlnValCysAsnSerGluGluCysProGlyThrTyrGlu 612
534 GCGGAGACCAAGCTGCAAGAAAGACCGCTGCC..... 570
612 spPheArgAlaGlnGlnCysAlaLysArgAsnSerTyrTyrValHisGln 628
570 ..... 570
629 AsnAlaLysHisSerTyrValProTyrGluProAspAsp..... 641
571 .....ATCAATGAGAGCGGTGCTGTGTCACCATGACATGTGTCTGT 617
642 .....AspAlaGlnLysCysGluLeuIleCysGlnSerAla. 653
618 CACCTGTGAGAGAGGGGTACAGAAAGTATGCTGTCTGCTGCACTGA 667
654 .....AspThrGlyAspValAlaPheMetAsnGlnVal..... 664
668 GAATGACTGAGAGCAAGAGAGTGGCCCAATGAGCTGAGGCGGCTGCC 717
665 .....ValHisAspGlyThrArgCysSerTyrArgAspProTyrSerVa 679
718 CTATGCTATCAGACAGAGAGTTCAGTACAGAAATTAACAGAGATGAGCT 767
679 LysAlaArgGlyGluCysValProValGluLysAspLysGluValGlys 696
768 T..... 768
696 erMetLysAlaAspAspLysCysGlyValCysGlyLysAspAsnSerHis 712

```

```

seq_name:/cgn2_6/ptodata/2/lna/6a_COMB.seq.us-08-706-216-5
seq_documentation_block:
  Sequence 5, Application US/08706216
  Patent No. 6140098
  GENERAL INFORMATION:
    APPLICANT: Balaubramanian, Sriram
    APPLICANT: Ford, John
    APPLICANT: Gorman, Daniel M.
    APPLICANT: Zurawski, Gerard
    TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS
    NUMBER OF SEQUENCES: 6
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: DNAX Research Institute
      STREET: 901 California Avenue
      CITY: Palo Alto
      STATE: California
      COUNTRY: USA
      ZIP: 94304-1104
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/706,216
      FILING DATE: 30-AUG-1996
      CLASSIFICATION: 435
      ATTORNEY/AGENT INFORMATION:
        NAME: Ching, Edwin P.
        REGISTRATION NUMBER: 34,090
        REFERENCE/DOCKET NUMBER: DX0613
        TELECOMMUNICATION INFORMATION:
          TELEPHONE: 415-852-9196
          TELEFAX: 415-496-1200
      INFORMATION FOR SEQ ID NO: 5:
        SEQUENCE CHARACTERISTICS:
          LENGTH: 1173 base pairs
          TYPE: nucleic acid
          STRANDEDNESS: single
          TOPOLOGY: linear
        MOLECULE TYPE: cDNA
        FEATURE:
          NAME/KEY: CDS
          LOCATION: 1..1173
        US-08-706-216-5
alignment_scores:
  Quality: 293.00      Length: 407
  Ratio: 1.363         Gaps: 21
  Percent Similarity: 52.826   Percent Identity: 29.238
alignment_block:
  US-09-938-330-20 x US-08-706-216-5 ..
Align seg 1/1 to: US-08-706-216-5 from: 1 to: 1173
  76 LeuPRArghISserSerHISleuArGValAlaArgSer...ProLeuH1 91
      |||||||
      |||||||
  97 TTGCGGGGGTGTGCGCAGAGCCGCCGAGTCCGCGGAGCTCGCCCTCCA 146
      |||||||
      |||||||
  91 sPrGIGlYlThreutPrProGlyArGValGlyArghISserLeuTyP 108
      |||
      |||
  147 CTTGGGC.....CTGCACCCAGAGAGGGGTGAGC.....T 175
      |||
      |||
  108 heAsuValTrhValPheGlyLySglLeuHISleuArGleuArGpRoxAs 124
      :::::
      :::::
  176 AGCGCTTGGGGCCACAGGGCAACAATTCAACCTCCACCTCGGGAAGAAC 225
      :::::
      :::::
  125 ArgArGleuValAlaProGlyISserValGluTyP...GInGluAspPh 140

```





9201 TCACCCAGCGCGTCCCGTACCGTGGCGCTCCACACATCC..... 9160  
 52 LeuSerHisValValSerGlyProAlaIleSerAlaGly\_SerMet 68  
 9159 ...ACGTCGGGGGACACCGCCGATCGACGAGCGCTCACGGA... 9117  
 69 ValValAspThrProProThrLeuProArgHisSerSer\_HisLeuArg 85  
 9116 .....TCACCGCGTGGCGCGGACCACTCGCGCGCTGAGCCCA 9076  
 85 AlaIleArgSerProLeuHisProGlyGlyThrLeuTrpProGlyArgVal 101  
 9075 TTGCTCCG.....ACCGTCGATTGACCGGCACT 9047  
 102 GlyArgHisSerLeuTyrPheAsnValThrVal.....PheGly 115  
 9046 GCCCGGACCGACCGCCGACCGCCGATACCGGTGGCGCGCGCTCGGAAA 8997  
 115 sGluLeuHisLeuArg.....LeuArgProAsnArgArgLeu 128  
 8996 GCCGCTCCACCGACGACACCCACACCTCCGCCGACGCGCTCCGCTC 8948  
 128 AlaValProGlySerSerValGluTrpIleGlnAspPheArgGluLeuPhe 144  
 8947 .....CGACCGCTCCGCGAAGCCCTT 8927  
 145 ArgGlnProLeuArgGlnGlnCysValTyrThrGlyVal\_ThrGly 161  
 8926 GCACCGCGCGTCCGCGCGCCGACCGCGCTCCGCGAAGCTCCACGACAA 8877  
 161 eTrpGlyAlaIleValAlaIleSerAsnCysAspGlyLeuAlaGlyLeu 177  
 8876 CCGTGGCGCGTCCGATCACCCTACCC..... 8849  
 178 IleArgThrAspSerThrAspPheIleGluProLeuGluArg..... 192  
 8848 .....ACCGCCACAGCCAGATCACCTACCCCGCCGACGACTGCA 8805  
 193 GlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 207  
 8804 CCGCGACATGCAACCGCCACGACGACGACGACGACGACGACGACGACG 8755  
 208 ArgArgGlnAlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 222  
 8754 AGCGCGCGCGCTCCAAACCGCATGAGCAACCGCCGACGAGATCAC 8705  
 223 .....LeuHisAsnGlnAlaPheGlyLeuGlyAspLe 233  
 8704 ACTGCCGAGGCTCCCGTGAACATGCCCGCGTGAAGTGTGGCGGCT 8655  
 233 uProAsnLeu.....LeuGlyLeuValGlyAspGlnLeuGlyAspThr 247  
 8654 CCGCCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 8605  
 248 .....GluArgGlyArgArgHisAlaGlyProGlySerTyr 259  
 8604 GTCCGGGTCCCGCGCGAGTCCGGAACGATGCCGCGCG..... 8564  
 260 SerIleGluValLeu.....LeuValValAspSerValValArgPhe 274  
 8563 ...CTCCAGCGCTCCACGAGCTCTCCAGCAGCAACCGCTGGCGGT 8517  
 274 eHisGlyGlyGlnHisValGlnAsnTyrValLeuThrLeuMetAsnIle 291  
 8516 CCATGGCGCTGGCTCCGCGCGCATAT..... 8499  
 291 AlaAspIleTyrHisAspGlnSerLeuGlyValHisIleAsn..... 305  
 8488 .....CCGGAAGACTCCGGTGAACCTCGCGCGCGGA 8456  
 306 .....IleAlaLeuValArgLeuIleMetVal 314  
 8455 ACGCAGAACCGCGCTCCCGTACATACCTTCCGCGGACCGCGCGCT 8406

314 IcIlyArgIleSerLeuSerLeu.....IleGluArg 326  
 8405 CCGGCTCTACAGGCCCTCGACGATCCAGCCAGCGTGGCGGGAACGG 8356  
 326 LysAsn.....ProSerArgSerLeuGlnValCysArgTyr 338  
 8355 GAGACCGTGTCCCGTCAACCGCCACAGGTCCACAGTGTGTCCGCGCA 8306  
 339 AlaHisSerGlnGlnArgGlnAspProSerHisAlaGlnHisAspHis 355  
 8305 CGCCACAC...GCCCGGTACGCGACCGCCATGCCCATGCGATCG 8259  
 355 sValValPheLeuThrArgGlnAspPheGlyProSerGlyTyrAlaPro 372  
 8258 GTCACGCTGTCCGCGATGACCGCGCGTGGGGCGG..... 8222  
 372 aThrGlyMetCysHisProLeuArgSerCysAlaLeuAsnHisGluAsp 388  
 8221 .....CGTGGCGGCTCCGCGCTCCGCTTC 8198  
 389 GlyPheSerSerAlaPheValIle.....AlaHisGluThrG 401  
 8197 GGG...ACCGAACAGTTCGTGTGCAAGTGGAGCGGAGCGAGGGGG 8151  
 397 .....AlaHisGluThrG 401  
 8150 TCGGCTGTGAGAGCAGCGCTTCGCGAGTTACGCGCGGTGACGCG 8101  
 401 LysHisValLeuGlyMetGlnHisAspGlyGlnGly\_AsnGlyCysAlaAs 417  
 8100 GTCAGCGGCTTACCGAGTTGAGAGACGCTGAGGAGTGAACCGCGCA 8051  
 417 P...GluThrSerLeuGlySerValMetAlaPro.....Leu 429  
 8050 CCGGAAGCGCAGCTCCGGGTGACCGCGCTGCGCGCGGCGGCGGCGCA 8001  
 429 aGlnAlaAlaPheHisArgPheHisIleTrpSerArgCysSerLeuGlu 445  
 8000 TCCGCGCGCTCCGAGCGCACAAAGCGTGAAGACAGACGAGCGCTTC 7951  
 446 Leu.....SerArgTyrLeuProSer...TyrAspC 455  
 7950 GCGCGGAGGCGCGCGACGAGACACCGCGCGCTCCGCGCGCTCTG 7901  
 455 sLeuLeuAspAspProPheAspProAlaTrpProGlnProGluLeu 472  
 7900 T.....CCGCTCCCGCGCGGTGCGCGCTGCGCGCGCGCTCC 7863  
 472 rGlyIleAsnTyrSerMetAspGlnGlnCysArgPheAspPheGlySer 488  
 7862 CTGGGTG.....AGAGCGGACGCGTGTCC 7837  
 489 GlyTyrGlnThrCysLeuAlaPheArgThrPheGlu...ProCysGly 503  
 7836 GCGGCAATGCCCGCGACCGCCGACGCGCGGCGGAGACCGCTGAGAAAG 7787  
 504 .....GlnLeuTrpCysSerHisAspAsnProTyrPheCysL 517  
 7786 CGACAGTACGCGGCGGTACCGAACCGCCGACGAGGTCTCCAGTCCA 7737  
 517 yThrIlyGlyGlyProProLeuAspGlyThrGlyCysAlaPro\_GlyL 533  
 7736 CTTGGCGAGAGAGAGAGCTCTCGCGCGGTGTGTGCTCCGCGCA 7687  
 533 yTrpCysPheGlyGlyHisCysIleTrpLysSerProGlnGlnThrTyr 549  
 7686 GAGGTCCCGCGCGCGCCACCGACGCTGTGAGAGAGTCCGAGCGCGCTC 7637  
 550 GlyGlnAspGlyGlyTrpSerSerTrpThrLysPheGlySerCysSer 566  
 7636 GGGGTCCAGGGC.....CGTACGCGCTCCGCTTCGAAAGACTTCGCG 7593



```

6146 CGTCCGGGTGAACACCGTCGCCAGCGCGCTGAAACATCTCCGCTCC 6097
1043 ntrpvalproglinserylpro..... 1050
6096 GCACGGCCCCGGCGATCGCCGACACACCATCATCACCAGCCCC 6047
1051 .....LeuHisProIleAsnLysIleSerSerMetCysAlaIaIaLupro 1065
6046 CAGGCACTACACCGCATCCCGCTCAGCAACCCATCCGAGACCCTCTCA 5997
1066 CysThrGlySparSerValIlePheCysGlnMetGluValLeuAspArgTy 1082
5996 CCAAGCTCCCAACACCCAGCAACCAACCCCAACCCCAATCCCGA 5947
1082 rcysserileprogltyrHisArgLeucCysValSerCysIleLysL 1099
5946 CGACACCGCCCTCCGCGCAT..... 5928
1099 ysAlaserGlyProAsnProGlyProAspProGlyProThrSerLeuPro 1115
5927 .....GACGACCCCAATACCCGCGATCCCGCAACTCCCGCGAACCACCA 5883
1116 ProPheSerThrProGlySerProLeuProGlyProGlnAspPro.... 1130
5882 CCCCACCCGACACCCGCGACACCCACCCGCGCGCGCGACCGACCG 5833
1131 ....AlaAspAlaIaLuproProGlyLysPro..... 1140
5832 AACTCCACGACTCCACACCCCGCGGAACTCCCGCAACACCGGCTCAT 5783
1141 ..ThrGlySerGluAspHis...GlnHisGlyArgAlaThrGlnLeuPro 1155
5782 CAACACCGGATGAACCCCATGGACACATCCAAACCGCGCGACATCCACC 5733
1156 GluAlaLeuAspThrSerSerProGlyThr...GlnHisProPheIaLp 1171
5732 CATCCCGCATCAGCAGGACACCGCATCCGACACCCACATCATCC 5683
1171 ogLutThrProIleProGlyAlaSerTrpSerIleSerProThrThrPro 1188
5682 GACACACACCGACCGCGCGCATTCACCGCGCAACCGACACCCACTC 5633
1188 LysGlyLeuProTrpGlyTrpThrGlnThrProThrProValProGlnAsp 1204
5632 CCCCACCCCTTACACACCCCGCGACACCGACTCCGACCCCGCCACCG 5583
1205 LysGlyGlnProGlyGluAspLeuArgHisProGlyThrSerLeuProAl 1221
5582 ACCACATCCCCC.....CACCCACCGCGCAACCCACCCCATC 5548
1221 AspLeuProGlyArgProPro...GluProCysHisProThrGlyThr 1236
5547 AACCCACCCCGCGCGCACACCGCAACCGCATCCCAACGACACCA 5499

```

seq\_name: /cgn2\_6/prodata/2/lna/5B\_COMB.seq:US-08-804-227C-7

seq\_documentation\_block:

```

; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostek, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4437 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-227C-7

```

alignment\_scores:

Quality	280.50	Length	1454
Ratio	0.467	Gaps	80
Percent Similarity	41.334	Percent Identity	21.802

alignment\_block:

US-09-938-330-20 x US-08-804-227C-7/rev ..

Align seg 1/1 to reverse of: US-08-804-227C-7 from: 1 to: 4437

```

36 SerAspGlyGluValThrValProCysSerThrAspPheArgLysArg 52
9216 TCACCGACGCGCGTCCCGGTACGCGCGCTGCCACACCATCC..... 9175
52 eLeuSerHisValAlaSerGlyProAlaAlaAlaSerAlaGly.SerMet 68
9174 ....ACGTGGGGGGGCAACAGCCCGCATCAGCCAGGCTCTCAAGGA... 9132
69 ValValAspThrProThrIleLeuProArgHisSerSer.HisLeuArg 85
9131 .....TCACCGCGTGGTCCCGCGACCACTCGCGCGCGCTGACGACCA 9091
85 AlaAlaArgSerProLeuHisProGlyGlyThrIleuTrpProGlyAlaGlu 101
9090 TTGCTCCG.....ACGCTCTGATTCACGCGACT 9062
102 GlyArgHisSerLeuTyTrpPheAsnValThrVal.....PheGlyTy 115
9061 GCGCGGACACGACGCGACGACCGCATGACGCGTGGCGCGCGGTGGGAAA 9012
115 sGluLeuHisLeuArg.....LeuArgProAlaGluArgLeu 128
9011 GCGGCTCCACGACGACGACGACGACGACGACGACGCGCGCTCCGCTC. 8963

```

[illegible]

389	glyPheSerSerAlaPheValIle.....	396
8212	GGG.....ACCGAACAAGTTCGTCTCAGAGTGGAGCCGACGAGAGGGGG	8166
397	.....AlaHisGluThrG	401
8165	TCGGGTGGTCCGAAGACAGAGCGTGTCCGACAGTTCACGTCGCGCGTGGACG	8116
401	LYHISValLeuGLyMetClnHisAspGlyGlnGly..AsnGlyCysAlaAs	417
8115	GTACAGCGGTTACGAGATTGACAGAAACGGTGGAGGATGCAACCGGGGGA	8066
417	P.....GluThrSerLeuGlySerValMetAlaPro.....LeuV	429
8065	CCGGAAGCCGACACTCCGGGTGCAGCGCGTGGCCGACGAGCGTGGCGCAGGA	8016
429	aIcGlnAlaAlaPheHisArgPheHisIleTPSerArgCysSerLyLeuGlu	445
8015	TCGCCGCGCGCTCCGAGCGACACACCGTGCAGACAGCAGCAGCGCTTCC	7966
446	Leu.....SerArgLyLeuProSer...TyrAspCy	455
7965	GCCGCGAGAGGGCCGACGGAAGACCGACGCGCGCTTGGCCCCCGTCTCC	7916
455	sLeuLeuAspAspProPheAspProAlaIleTPProGlnProProIleuP	472
7915	T.....CCGCTCCCGCCGGGCGCGCGCTGCGCGCGCGCGCTCC	7878
472	roGlyLeuAsnTyIleSerMetAspGlnGlnCysArgPheAspPheGlySer	488
7877	CTGGGGTG.....AGACGCGCAGCGGTGCTGC	7852
489	gLYTyrgLnThrCysLeuAlaPheArgThrPheGlu...ProCysLys..	503
7851	GCGGCATGGCGCCGACACGCGCCGACAGCGGGCGGAGACCGCTGCAAGAG	7802
504	.....GlnLeuTrpCysSerHisProAspAsnProIlyrPheCysL	517
7801	CGACCAAGTACGGCGGGTACCGAACCAGGCCCGACGAAGTCTCCAGTCCA	7752
517	ySThrLysLysGlyProProLeuAspGlyThrGlyCysAlaPro..GlyL	533
7751	CGTCCGGGACAGCAGCAGAGAGCTCTCCGCGCGCTCGTCTCCCTCCGCGCA	7702
533	ySTrPrcysPheLysGlyHisCysIleTPrlySerProGlnGlnThrTyI	549
7701	GAGGTCCCGCGCGCCACACAGCTGTCTGAAGAGCTGAGGCGCGCTC	7652
550	GLyGlnAspGlyGlyThrPheSerIleThrLysPheGlySerCysSerArg	566
7651	GGGGTCCAGGCG.....CGTACGCGCGTCCGTTGGAAGGACTTCGCGG	7608
566	gSerCysGlyGlyGlyValArgSerArgSerArg.....	577
7607	CCGTTCCCGAGGGCGATTCGCGCACCGGCCACAGGCCCCCAAGCGATGAG	7558
578	..SerCysAsnAsnProSerProAla.....	585
7557	GCGGCACGGGCTCCCTCGCGCCGCGCCGCTGCGCAAGGGCGTCAAGGC	7508
586	..TyrGlyLysArg.....ProCysLeuGlyProMetPheG	597
7507	GCGGTGCGGCGCGGACGAGCGCGCTCGCTCCGTTGGCCCA.....	7466
597	uTyrgLnValCysAsnSerGlnGlnCysProGlyThrTyIrgLysAspPheA	614
7465	.CAAGCGCAGCATGAGAGAAAGACACGAAGCGTCCGACGCCGAAGTCC	7411
614	rgaGlnGlnGlnCysAlaLysArgAsnSerTyIryValHisGlnAsnAla	630
7416	CGCGGTCAAG.....	7409
631	LYHISserTrpValProTyIrgLupProAspAspAlaGlnLysCysel	647

7409 ..... 7409  
647 uLeuIleCysGlnSerAlaSprThrIlyVal ..... 659  
7408 .....CTGCTCCAGGTAGCGGGCGGCTGCTCCAGTACGGCG 7363  
660 .....PheMetAsnGlnValValHisAspGly 668  
7362 TCACCCGGCTGGAGGTGAGCGAGTGCAGACACACCGTCTGTCGAGCAACC 7313  
669 ThrArgCysSerTyrArgAspProTyrSerValCysAla ..... 681  
7312 TGCCGGTCCAGTACGCGCAGCAGGCGCGCTCTCCGGGAGGCCCGGA 7263  
682 .....ArgGlyGluCysValProVal .....GlyCysAspGly 693  
7262 GAGCGCGGCGCGCTGCTCCGCTCCGCGACGTCGACGCGCTCCAGCGGC 7213  
693 LuValGlySerMetLysAlaAsp .....AspLysCysGlyValCysGly 707  
7212 ACTTGGAAACGAGCGCGCTCCAGTCTCCGACGAGTCCGCTCCGCG 7163  
708 GlyAspAsn .....SerHisCysArgThrValLysGlyThrLeuG1 721  
7162 GCG 7119  
721 LysAlaSerLysGlnAlaGlyAlaLeuLysLeuVal ..... 733  
7118 CCGACCG 7069  
734 .....GlnIleProAlaGly .....AlaArgHisIleGln 743  
7068 CCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7019  
744 IleGluAlaLeuGluLysSerProHisArgSerValLysAsnGln 760  
7018 GTCCAGAGGTGCTGCTCCGCTCCGCTCCGCGCGCGCGCGCGCGCGCGCG 6969  
760 a1 .....ThrGlySerPheIleLeuAsnPro 768  
6968 TGGTACGCTGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6934  
769 LysGlyLysGluAlaThrSerArgThrPheThrAlaMetGly .....Le 783  
6933 ...GGCGGGGAGACACACGCGCGACATGACGCGCGAGCGCGCTACCG 6887  
783 uGluTrp ..... 785  
6886 GAACTGCTTCTCTCCCGCGGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCA 6837  
786 .....GluAspAlaVal .....GluAspAlaVal ..... 793  
6836 CCGCGGGGAGCGACCTCCCGGGCAGTCCGCAAGCGCGCGCGCGCGCGCGCG 6787  
794 ...GluSerLeuLysThrSerGlyProLeuProGluAlaIleAlaIle 809  
6786 CCCAGTTCAGCGCGGACCTGTCGAGTCCCGACAGTACGCG ..... 6742  
809 uAlaLeuProThrGluGly .....GlyProArgSerSerLeuAla 824  
6741 .....CCCCGAGGAGAGGGGTCTGCTCCGCGCGCGCGCGCGCGCGCGCT 6699  
824 yTrpLysTyrValIleHisGluAspLeuLeuProLeuIleGlySerAsnAsn 840  
6698 CACGGGTGGTATCCAGAGC .....GGGCTTCGGGG 6667  
841 ValLeuLeuGluGluMetAspThrTyrGluTrpAlaLeuLysSerTrp ... 856  
6666 ACCGCC .....TCGCGCGAGCGCGCTGGAG 6644  
857 ...AlaProCysSerLysAlaCysGlyGlyGlyIleGlnPheThrLys 871  
643 GAGGGCAGCGCGCTCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGAGAA 6594  
872 TyrGlyCysArgArgArgArgArgArgArgArgArgArgArgArgArgArg 888  
6593 CTGGGGTGGTGGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6567  
888 AspHisLysLysArgPro ..... 894  
6566 .....CG 6527  
895 ...LysProIleArgArgArgArgArgArgArgArgArgArgArgArgArgArg 910  
6526 ACCGACCG 6477  
911 TrpValThrGluGluTrpGlyValLysSerArgSerCysGlyLysLeuG1 927  
6476 CAGTCTCGATCGCGGGGCGGAGTCCGAGTCCGCGCGCGCGCGCGCGCGCG 6427  
927 yVal .....GlnThrArgGlyIleGlnCysLeuLeuP 938  
6426 GCGAGAGCGACACATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6387  
938 roLeuSerAsnGlyThrHisLysValMetProAlaValAlaCysAlaGly 954  
6386 .....GTCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6348  
955 AspArgProGluAlaArgArgProCysLeuArgValProCysProAlaG1 971  
6347 TCCGCGCGCGCTGTCG .....ACCGTCCCGCTCA 6319  
971 rTrp .....A 973  
6318 GTGGCG 6269  
973 rGleuGlyAlaTrpSerGlnCysSerAlaThrCysGlyGlyIleGln 989  
6268 GTGCGGAAAGGGGTCCAGCGAGTACCGGTCCGCTGGAAAGGGT ..... 6225  
990 GlnArgGlnValValCysArgThrAsnAlaAsnSerLeuGlnHisCysG1 1006  
6224 .....AGCTGGGAGGTGATGATGATGATGATGATGATGATGATGATGAT 6193  
1006 uGlyAspArgProAspThrValGlnValCysSerLeuProAlaCysGly 1023  
6192 CCGG .....TGGGAGTCCG ..... 6177  
1023 LysAsnHisGlnAsnSerThrValArgAlaAspValTrpGluLeuGly ... 1038  
6176 .....TGGCTCCAGGCCG 6162  
1039 .....ThrProGluGly 1043  
6161 CGTCCCGGGTGACACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6112  
1043 nTrpValProGlnSerGlyPro ..... 1050  
6111 GCAGCG 6062  
1051 .....LeuHisProIleAsnLysLysSerSerMetCysAlaIleGluPro 1065  
6061 CAGGCACTACCGCGCATCCCGTCCAGACCCCATCGGAGCCACCTCA 6012  
1066 CysThrGlyAspArgSerValPheCysGlnMetGlnValLeuAspArgGly 1082  
6011 CCAAGCTCCACACCGACGACGACGACGACGACGACGACGACGACGACGACG 5962  
1082 rCysSerLysProGlyLysHisArgLeuCysCysValSerCysLysLysL 1099  
5961 CGCACCGCTCCCGCGCAT ..... 5943  
1099 yAlaSerGlyProAsnProGlyProAspProGlyProThrSerLeuPro 1115  
5942 .....GAGGACCCAAATACCCCGATCCCAATCCCGCGCGCGCGCGCGCGCG 5898

```

193 .GlyGlnGlnIuLysGluAlaSerGlyArgThrHisValTyr... 207
      |||:::||||::: |||:::
8819 CGCCAGATGCAACGCCACCACGAGCAGACGACACGCGCTGTCCACCGTC 8777

```

```

208 ..ArgArgGluAlaValGlnGlnIleuTrpAlaGluProAspGlyasp... 222
      |||||
      |||||
8769 AGGCGCGGCGCCCTCCAAACCCCAACGATAGACCCGACGCGGAGATAC 8720
223 .....LeuHisAsnGluAlaPheGlyLeuGlyAspLe 233
8719 ACTGCGAGGATCCCGGTGACATAGCCCGGTAGGTGTGGGGGT 8670
233 uProAsnIleu.....LeuGlyLeuValGlyAspGlnLeuGlyAspThr. 247
      |||||
      |||||
8669 CCCCAGGTGGGTGCGGTAGTCTGTGTGGAGATCCCGTGAACACACG 8620
248 .....GluArgLysArgHisAlaLysProGlySerTyr 259
      |||||
      |||||
8619 GTCCGGGTGCGCGGACGAGTCGGAACGATGCCGCGCGC..... 8579
260 SerIleGluValLeu.....LeuValAlaAspAspSerValAlaArgPh 274
      |||||
      |||||
8578 ...CTCCAGCGGCTCCCAAGACGTCTCCAGACGAAACCGGTGTGGGT 8532
274 eHisGlyLysGlnHisValGlnAsnTyrValLeuThrIleuMetAsnIleV 291
      |||||
      |||||
8531 CCATGGCCCTGGCTCGCGCGGATAT..... 8504
291 aAlaAspGlnIleTyrHisAspGlnSerLeuGlyValHisIleAsn..... 305
8503 .....CCGGAAGAACTCCGCGTGAACCTCGCGCGCGCA 8471
306 .....IleAlaLeuValAlaGlyLeuIleMetVa 314
8470 ACCGAGAAACCGGCTCCCTCCCTACATAGCTGTGCGCGCAACCGCGCT 8421
314 lGlyTyrArgIleuSerLeuSerIleu.....IleGluArg 326
      |||||
      |||||
8420 CCGGGTGTGACAGCCCTCGACGTCCAGCCAGCGGTGCGCGGAGAGGG 8371
326 Lysn.....ProSerArgSerLeuGlnValLysArgTyr 338
8370 GAGACCGGTGCGCGCTCAACGCGCACAGTCCCAAGGTGTGCGCGCA 8321
339 AlaHisSerGlnArgGlnAspProSerHisAlaGlnHisAspHis 355
      |||||
      |||||
8320 CCGCCACAC...GCCGGGTAAAGGACACGACGCGATCCCAAGTGTGCGCGCA 8274
355 sValValPheLeuThrArgGlnAspPheGlyProSerGlyTyrAlaProV 372
      |||||
      |||||
8273 GCTACGCTGTGCGCGCATGACGCGGTGGGGGGCGC..... 8237
372 aThrGlyMetCysHisProLeuArgSerCysAlaLeuAsnHisGluAsp 388
8236 .....CGTGCCTGGCTCGCGCTGCTGTC 8213
389 GlyPheSerSerAlaPheValIle..... 396
      |||||
      |||||
8212 GGG...ACCGAACAAGTGTGTGTGTGACAGTGGAGGCGAGCGAGGGGG 8166
397 .....AlaHisGluThrG 401
8165 TCGGGGTGTGAGAGAGCGGTGTTCGCGAGTTCAGAGCCCGTGGAGCG 8116
401 LysHisValLeuGlyMetGlnHisAspGlyGlnGly..AsnGlyCysAlaAs 417
      |||||
      |||||
8115 GTCCAGGCGGTACGACGATTCGAGAACGTTGAGAGTGCAGAACCCGCGGA 8066
417 P...GluThrSerLeuGlySerValMetAlaPro.....LeuV 429
      |||||
      |||||
8065 CCGGAAGGCGAGGCTCGGGGTGACAGCGGTGCGCGGAGCGCTGGCGAGA 8016
429 aGlnAlaAlaPheHisArgPheHisTrpSerArgCysSerIleuGlu 445
      |||||
      |||||
8015 TCCCGGCGGCTCGAGCGCAAGCGGTAGACAGACGACGCGCGCTTCC 7966

```

---

```

446 Leu.....SerArgTyrIleuProSer...TyrAspCy 455
      |||||
      |||||
7965 GCCGCGAGGAGGCGCAGGAGAGCACGCGCGCGCTGCGCGCGCTGCTG 7916
455 sLeuLeuAspAspProPheAspProAlaTrpProGlnProProGluLeuP 472
      |||||
      |||||
7915 T.....CGGCTCCGCGCGGTGCGCGCGCGCGCGCGCGCGCTCC 7878
472 rGlyIleAsnTyrSerMetAspGlnGlnCysArgPheAspPheGlySer 488
      |||||
      |||||
7877 CTGGGCTG.....AGAAGCGCACGGTGTCTCC 7852
489 GlyTyrGlnThrCysLeuAlaPheArgThrPheGlu...ProCysLys.. 503
      |||||
      |||||
7851 GCGGATGCGCGCGCACGCGCGCGGAGAGCGCGCGCGCGCGCGCGTCA 7802
504 .....GlnLeuTyrCysSerHisProAspAsnProTyrPheCysL 517
7801 CGACCGATGCGCGCGGTGACCGAGAGCGCGCGCGCGCGCGCTCCCA 7752
517 yThrIleLysGlyProProLeuAspGlyThrGlyCysAlaPro..GlyL 533
      |||||
      |||||
7751 CCGCGCGCAGCAGAGGAGCTCTGCGCGCGCGGTGCGCTGCGCGCA 7702
533 yThrProCysPheLysGlyHisCysIleTyrPlySerProGlnGlnThrTyr 549
      |||||
      |||||
7701 GAGGTCCGCGCGCGCGCACACGTCTGAGAGAGGTGACGCGCGCGCTC 7652
550 GlnGlnAspGlyGlyTyrPheSerThrPheLysPheGlySerCysSerAr 566
      |||||
      |||||
7651 GGGGTCCAGGCG...CGTACCGCGCTCCCGTTCAGAGACTTCGCGC 7608
566 gSerCysGlyGlyValAlaArgSerArgSerArg..... 577
7607 CGGTTCGAGGCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGATCGAG 7558
578 ..SerCysAsnAsnProSerProAla..... 585
7557 GCGGACGCGGCTCCCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7508
586 ..TyrGlyLysArg.....ProCysLeuGlyProMetPheG 597
      |||||
      |||||
7507 GCGGTGGCGCGCGGTGACGAGCGCTGCGCGCGCGCGCGCGCGCGCGCG 7466
597 uTyrGlnValCysAsnSerGlnGlyCysProGlyThrTyrGlnAspPhea 614
      |||||
      |||||
7465 ..CAGCGCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7417
614 rGlnGlnGlnCysAlaLysArgAsnSerTyrTyrValHisGlnAsnAla 630
      |||||
      |||||
7416 GCGGTGAC..... 7409
631 LysHisSerTrpValProTyrGluProAspAspAlaGlnLysCysGl 647
7409 ..... 7409
647 uLeuIleCysGlnSerAlaAspThrGlyAspValVal..... 659
7408 ...CTGCTCAGGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7363
660 .....PheMetAsnGlnValValHisAspGly 668
7362 TCCACCGCGGTGAGAGTGCAGAGTGCAGACACGCTGCTCGAGCACAC 7313
669 ThrArgCysSerTyrArgAspProTyrSerValCysAla..... 681
      |||||
      |||||
7312 TCCGCGCTGACGAGCGCGAGAGGAGCGCGCTGCGGAGAGCGCGCA 7263
682 ....ArgGlyCysValProVal.....GlyCysAspLysG 693
      |||||
      |||||
7262 GAGAGCGGCGAGTGTGCTCCGCTGCGCGAGCTGCGAGCGCGCTGCGAG 7213
693 LuValGlySerMetLysAlaAsp.....AspLysCysGlyValCysGly 707

```





1188 1yg1yLeuProTrrp1ytrThrInThrProthProValProGluasp 1204  
1189 |||||  
5647 CCCCAACCCCTACACACACCCCGCCACACCGACCTCCAGCCGCCACCC 5598  
1205 1y5g1yGlnProG1yGluaspLeuArgHisProG1yThrSerLeuProAl 1221  
1206 |||||  
5597 ACCACATCCCCC.....CACCACCGGACACCCACCCATC 5563  
1221 aasPleuProG1yArgProPro...GluProCysHisProthG1yThr 1236  
1222 |||||  
5562 AACCCACCCCGCCGACACACCGACCGCATCACCACACGACACACA 5514  
seq\_name: /cgn2\_6/plodata/2/ina/6h\_comb\_seq:us-09-351-414-1  
documentation\_block:  
Sequence 1, Application US/09351414  
Accession No. 6265199  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
APPLICANT: Baindur, Nand  
APPLICANT: Deisher, Theresa A.  
APPLICANT: Bishop, Paul D.  
TITLE OF INVENTION: DISINTEGRIN HOMOLOG  
FILE REFERENCE: 98-29  
CURRENT APPLICATION NUMBER: US/09/351,414  
CURRENT FILING DATE: 1999-07-09  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 2268  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3)...(2090)  
NAME/KEY: misc\_feature  
LOCATION: (1)...(2268)  
OTHER INFORMATION: n = A,T,C or G  
US-09-351-414-1  
alignment\_scores:  
Quality: 272.00 Length: 698  
Ratio: 0.842 Gaps: 35  
Percent Similarity: 46.275 Percent Identity: 21.347  
ment\_block:  
09-938-330-20 x US-09-351-414-1 ..  
Align seg 1/1 to: US-09-351-414-1 from: 1 to: 2268  
108 pheasnValThrValPheG1yLysG1yLeuHisLeuArgLeuArgProAs 124  
109 |||||  
90 TTCAGATTGAAGCTTCGGCTCCAAATTCATCTTGACCTCCTACTGTA 139  
124 naGArgLeuValValProG1ySerSerValG1uTrpGlnGluaspPhe 141  
125 |||||  
140 CAATGGCTTGTG.....TCTCTGATTATG 165  
141 rglLeuPheArgGlnProLeuArg.....Gln 150  
142 |||||  
166 TGGAGATTCTACTGAAATGGAAACACACACTCTAAGGCTGAGAG 215  
151 GlucysValIytrThrG1yValThrG1yMetProG1yAlaAlaValAl 167  
152 |||||  
216 CAGTCTACTACATGGAAGCATCAGAGCGTCAAGACTCCAGAGTGGC 265  
167 allelesAsnCysAspG1yLeuAlaIyLeuIleArgThrAspSerThra 184  
168 |||||  
266 TCTGTCAACCTGCATGACCTTCATGCGATGTTTGAAGATGATACCTTCG 315  
184 spPhePheIleGluProLeuGluArgG1yGlnGlnGluaspLeuAsp 200

316 TGTATATGATAGAGCCACTAGAGCTGTTCATGATGAGAAA...AGCA 362  
317 |||||  
201 GluArgThrHisValValIytrArgG1uAlaValGlnGlnGluTrpAl 217  
202 |||||  
363 GGTTCACACCATATATATCCAGAAAACCTTGGCAGACAGTATTTT... 407  
217 aGluProaspG1yAspLeuHisAsnGluAlaPheG1yLeuG1yaspLeu 234  
218 |||||  
408 .....AAGCAATGA 417  
234 roAsnLeuGlnG1yLeuValG1yAspGln.....Leu 244  
235 |||||  
418 AGAATCTCATATGGAAGAGGTACCAAGTGGCCCTTCTCTGTGAATTA 467  
245 GluAspThrGluArgIyArgArgHisAlaIyPro.....G1ySerTy 259  
246 |||||  
468 CAGTGGTTGAAAAGAGAGAGAGAGAGTGAATCATCAGCTGTATAT 517  
259 rSer.....IleGluValLeuLeuValIyAspAspSerValY 272  
260 |||||  
518 TGAAGAAATGAATATTTGAACTTATGATTTGTAATGACAAACGT 567  
272 alArgPheHis.....G1yLysG1uHisValGlnAsnTytrValLeuThr 286  
273 |||||  
568 ATAAACACATCGCTCTTCTCATCATACACACACCTTGGAAAGTCC 617  
287 LeuMetAsnIleValIyAspG1uIleTytrHisAspIySerLeuG1yAla 303  
288 |||||  
618 GTGGCAACCTTGTGATTTATTTACAGAGAGAG..... 653  
303 sIleAsnIleAlaIyLeuValArgLeuIleMetValG1yTytrArgGlnSer 320  
304 |||||  
654 .CTCAACACACAGGTTCTCTGTGGCTGTAGACACTGTGACTGAGAAAG 702  
320 euserLeuIleGluArgG1yAsnProSerArgSerLeuGlnGlnValCys 336  
703 ATCAGATTGACATCACCACACCCCTGTGCAGATGCTCCATGAGTTCTGA 752  
337 ArgThrAlaHisSerGlnGlnArgGlnAspProSerHisAlaGlnHis 353  
338 |||||  
753 AAATAC.....CGCAGAGCC.....ATTAAGCAGCATGC 781  
353 sasPHisValValPheLeuThrArgGlnAspPhe.....GlyProSerG 368  
782 TGATGCTGTGACCTCATCTCGGCTGACATTTCACTAAGAGAGCA 831  
368 IyTytrAlaProValThrG1yMetCysHisProLeuArgSerCysAlaLeu 384  
369 |||||  
832 GTCTGACTTACTTTGAAGGTGTCTGTCTCCGCAAGAGAGAGTGTGTG 881  
385 AsnHisGluaspG1yPheSerSerAlaPhe.....ValIleAlaHisG1 399  
386 |||||  
882 AAT...GAGTATGTCTCTCCAAATGGCAGTGGCAAGAGTATATCCAGAG 928  
399 uThrG1yHisValLeuG1yMetGlnHisAspG1yGlnG1y..... 412  
929 CTTGGCTCAAAACCTTGAATTCAAATGGGAACCTTTACCAAGAACCA 978  
413 .....AsnG1yCys...AlaAspG1uThrSer 420  
979 AATGTGACTGCACAGAAATCCTGGGTGGTGCATCATGAGAGAAACAGG 1028  
421 LeuG1ySerValMetAlaProLeuValGlnAlaAlaPheHisArgPhe 437  
1029 GGTCTC.....CATTTCTGAAA 1045  
437 sTrpSerArgCysSerTytrLeuG1uLeuSerArgTytrLeuPro.....S 452  
1046 ATTTTCAAGTGCAGCATTTTGGAGATATAGACATTTTACAGAGAGAG 1095  
452 eTytrAspCysLeuLeuAspAspPro.....PheAspProAlaTrp 465  
453 |||||

```

1096 GTGAGAGCTGCTTTTCAACAGGCAACAACTATTGAGCCACG... 1142
466 ProGlnProProGlnLeuProGlyLeuAsnYrSerMetAspGlnGly 482
1143 .....GAAATG 1147
482 sArgPheAspPheGlySerGlyTyrGlnThrCysLeuAlaPheArgThr 499
1148 T.....GAAATGATAC.....G 1161
499 heGluProCysGlnLeuThrProCysSerHisProAspAsnProTyrPhe 515
1162 TGGAAAGCTGGGAGAGGTGATGTGGTTTCATGCTGGAATGCTATGA 1211
516 CysLysThrLysGlyProProLeuAspGlyThrGluCysAlaProGln 532
1212 TTATGCTGTAAAGAAATGTTCCCTCCCAACGGGGCTCACTGACGACGG 1261
532 YLSTPRCysPheLysGlyHisCysIleTrp..... 542
1262 GCCCTGCTGTAAACATCTCATGCTCTTTTTCAGCCACGAGGGTATGAAT 1311
542 ..... 542
1312 GCCGGATGCTGTACAGAGTGTGATATTCATATATTTCTACTGAGAC 1361
543 .....LysSerProGlnGlnThrTyrGlyGlnAspGlyGlyTrpSerse 557
1362 TCTGTGATGCTCCCAACCAATCTTCATTAAGCAGACGGA..... 1400
557 rTrpThrLysPheLysSerArgSerCysGly.....G 570
1401 .....TATGATGCAATCAAAATCAGGGCCGCTGTACAAG 1437
570 LysGlyValArgSerArgSerArgSerCysAsnAsnProSerProAlaIyr 586
1438 GCGAGTGCAGAACCCAGACACACAGTGTGATAC.....ATCTGG 1478
587 GlyGlyArgProGlyLeuGlyPheGlyThrGlnValCysAsnse 603
1479 GGAACAAAGGCTGCAAGGCTGTGACAAATTTGCTATTAAGTGAATAC 1528
603 rGlnGlyCysProGlyThrTyrGluAspPheArgAlaGlnGlnCysAlaL 620
1529 AGAA.....GCACTGAGAGGA.....ACTCGGGGA 1557
620 ysArgAsnSerTyrTyrValHisGlnAsnAlaLysHis..... 632
1558 AGGATGAGAGCCGCTGATTC...CAGTGCAGCAAAACATGATGTCTCTGT 1604
633 .....SerTrpValProTyrGlnProAspAspAlaGln 644
605 GGATTCCTTACTCTGTACCAATCTTACGTGAGCTCCACGTATGTGCAACT 1654
644 nLysCysGlnLeuIle..... 649
1655 TCAGGTGAGATCATTCACACTCTTCTACATCAAGCCGGGTGATG 1704
650 ..CysGlnSerAla.....AspThrGlyAspValAlaPhe 660
1705 ACTGAGTGTGCTCCATGATCTTATGATGATGATGATGATGATGATGAT 1754
661 MetAsnGlnValAlaHisAspGlyThrArgCysSerTyrArgAspProTyr 677
1755 .....GTAGAGATGAGAACGCCATGTGGC.....CCGTC 1783
677 rSerValCysAlaArgGlyGluCysValProValGlyCysAspLysGluV 694
1784 TATGATGTGTATGATCGAAGTGCCTCAAAATTCAGGCC...CTAATA 1830
694 aGlySerMetLysAlaAspAspLysCysGlyValCysGlyGly 708
1831 TGAGCAGCTGTCCACTCGATTCAGGGAAGGTAAAGTGTGTGGGC 1874

```

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:us-08-243-542-8
seq_documentation_block:
; Sequence 8, Application US/08243542
; Patent No. 552526
GENERAL INFORMATION:
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: EKI, MITSURU
TITLE OF INVENTION: MDC PROTEINS AND DNAs
TITLE OF INVENTION: ENCODING THE SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLYNN, THIEL, BOUPELL & TANIS P.C.
STREET: 2026 Rambling Road
CITY: Kalamazoo
STATE: Michigan
COUNTRY: USA
ZIP: 49008-1699
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
COMPUTER: IBM PC/XT/AT Compatible
OPERATING SYSTEM: MS-DOS 5.0
SOFTWARE: WordPerfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/243,542
FILING DATE:
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 5-136602
FILING DATE: 14 MAY 1993
APPLICATION NUMBER: JP 5-257455
FILING DATE: 22 SEPTEMBER 1993
APPLICATION NUMBER: JP 6-49904
FILING DATE: 23 FEBRUARY 1994
APPLICATION NUMBER: JP 6-73328
FILING DATE: 12 APRIL 1994
APPLICATION NUMBER: JP 6-84470
FILING DATE: 22 APRIL 1994
ATTORNEY/AGENT INFORMATION:
NAME: Teriyence F. Chapman
REGISTRATION NUMBER: 32 549
REFERENCE/DOCKET NUMBER: Furuya Case 1313
TELECOMMUNICATION INFORMATION:
TELEPHONE: (616) 381-5465
TELEFAX: (616) 381-1156
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 3183 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
LIBRARY: human fetal brain cDNA library
FEATURE:
NAME/KEY: 3' UTR
LOCATION: 2308..3183
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2307
US-08-243-542-8
alignment_scores:
Quality: 267.00 Length: 1196
Ratio: 0.560 Gaps: 66
Percent Similarity: 39.883 Percent Identity: 21.321
alignment_block:
US-09-938-330-20 x US-08-243-542-8 ..

```

Align seg 1/1 to: US-08-243-542-8 from: 1 to: 3183

```

59 GLYPROLAAlaAla.....SerAl 65
   |||||
76 GGTCCCTGGAGCTGTGCGATGGGGGGCTTACCCAGCTGGAGGGCC 125
   |||||
65 aGlySerMetValValAspThrProProThrLeuProArgHisSer 81
   |||||
126 AGGAGCCCTGAGTTCAGGGAACCCAGCTGTGTTAGGAGAGCTCCG 175
   |||||
82 .....HisLeuArgValAlaArgSerPro 89
   |||||
176 GGGGAGAGTCCGAAGCAGAGCTGACACAGAGGTCCGCCAGAGACCA 225
   |||||
90 LeuHisProGlyGlyThrLeuThrProGlyArgValGlyArgHisSerLe 106
   |||||
226 .....CCAGGGGGGGGG.....CTGTCCATGTGGCCAGGTGAGT.. 261
   |||||
106 uThrPheAsnValThrValPheGlyLysGlyLeuHisLeuArgLeuArg 123
   |||||
262 .....TTGCTCATCCAGCCTTCAACTCACTCACTCCCTGGAGCTGGAGC 307
   |||||
123 roAsnArgArgLeuValValProGlySerSerValGlyThrGlnGlyAsp 139
   |||||
308 TGAACCCACCTCTCTC.....TCCCTGCAATACGTGGAG... 342
   |||||
140 PheArgGlyLeuPheArgGlnProLeuArgGln..... 150
   |||||
343 ...GCCCATCTCAGCCGGAGGAGCAACCCAGCAGACCGGGCTGG 389
   |||||
151 ...GlyCysValThrThrGlyValThrGlyMetProGlyAlaAlaVal 166
   |||||
390 AGACACACTGTACTACCAAGGAGACTCCGGGGAGACCCGACACTCTTCG 439
   |||||
166 aAlaIleSerAsnCysAspGlyLeuAlaGlyLeuIleArgThrAspSer 182
   |||||
440 CCGCCCTCTCCACTGCCAGGGGCTGATGGGGCTTCTGTGATGGAGAC 489
   |||||
183 ThrAspPhePheIleGlyProLeuGlyArgGlyGlnGlnGlyGlnVal 199
   |||||
490 TTGACTTACATCGTGAAGCCCAAGAG..... 516
   |||||
199 aSerGlyArgThrHisValValThrArgArgGlyAlaValGlnGlnGly 216
   |||||
517 .....GTGGCTGAGACTT 529
   |||||
216 rPalagLupProAspGlyAspLeuHisAsnGlnAlaPheGlyLeuGlyAsp 232
   |||||
530 GGGGAGGCCCTCAG.....GGAGCC 549
   |||||
233 LeuProAsnLeuLeu.....GlyLeuValGlyAspGlnLeuGlyAs 246
   |||||
550 CTTCGCCACTCATTTACCGAGCCCTCTCTCCAGATCCCTCGGATG 599
   |||||
246 pThrGlu..... 248
   |||||
600 CAGGGAACCAAGCTGTTGGCTGTGCTGCCAGCTCCGCTCTCCAA 649
   |||||
249 .....ArgLysArgArgHisAlaLysProGlySerYrSer 260
   |||||
650 ACCGGCCGAGGCTGAGAGAGAAAGCAGGTCCGGCGGAGCCACTTACA 699
   |||||
261 TLeuGlu.....ValLeuLeuValValAspAspSerVal 271
   |||||
700 GTGCACAGTGAACCAAGTATGTGAGACTAATGTGATCAACGACACCA 749
   |||||
271 lValArgPheHisGlyLysGlnHisVal.....GlnAsnTYrValL 285
   |||||
750 GCTTTTGGACAGATGGAGAGCTGCTCTCCACAGCAACTTTCGCA 799
   |||||
285 euThrLeuMetAsnIleValAspGlyIleThrHisAspGlnSerLeuGly 301
   |||||
800 ACTCCGTGTGAACCTGGCCGATGTGATATACAGAGACAG..... 840
   |||||
302 ValHisIleAsnIleAlaLeuValArgLeuIleMetValGlyTrArgG 318
   |||||
841 .....CTCAACACTCGCATCTCTCTGTGCTTCATGGAACATATGGCAGA 884
   |||||
318 nSerLeuSerLeuIleGlyArgGlyAsnProSerArgSerLeuGlnVal 335
   |||||
885 TGGGAGCAAGATCCAGATGCGAGATGACCTCTCTGAGAGACCTGGCCGGC 934
   |||||
335 aLysArgTrpAlaHisSerGlnArgGlnAspProSerHisAlaGly 351
   |||||
935 TCATGCTCTAC...CGACGGAGAGGTCTGCTGAGCCAGTAATGCCACC 981
   |||||
352 HisHisAspHisValValPheLeuThrArgGln...AspPheGlyProse 367
   |||||
982 CAC.....CTCTTCTCGGGCAGGACTTTCAGAGCAGCAGACAG 1019
   |||||
367 rGlyTrpAlaProValThrGlyMetCysHisProLeuArgSerCysAla 384
   |||||
1020 CGGGGACACCTTACGTGGGGGCATA.....TGTCTCC 1051
   |||||
384 euAsnHisGlnAspGlyPheSerSer.....Ala 393
   |||||
1052 TGTCCCATGGCGGGGTGTGAACGATACGGCAACATGGGGGGATGGCC 1101
   |||||
394 PheValIleAlaHisGlyThrGlyHisValLeuGlyMet.....G 407
   |||||
1102 GTGACCTTGGCCAGACGCTGGAGACAGACCTGGGCTGATGTGAACA 1151
   |||||
407 uHisAspGlyGlnGlyAsnGlyCys...AlaAspGlyThrSerLeuGly 423
   |||||
1152 ACACCGAGAGCTCGGAGAGGAGCTCAAGTGTCCAGACATGTGCTGGCT 1201
   |||||
423 eValMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTrpSer 439
   |||||
1202 GCATCATGAGAGACATGTGGTTCATCTGCCCGCAAGTTC.....T 1245
   |||||
440 ArgCysSerLysLeuGlnLeuSerArgTYrLeuPro.....SerTYrAs 454
   |||||
1246 CGGTGCACATCGACAGATACACCAAGTTCTGACAGAGGTTGTGGCAG 1295
   |||||
454 pCysLeuLeuAspAspProPheAspProAlaTrpProGlnGly 471
   |||||
1296 CTGGCTCTTCAACAGAGCCCTCAAGCTCTCG.....GACCCCGCAGAT 1339
   |||||
471 euProGlyIleAsnTYrSerMetAspGlnGlnCysArgPheAspPheGly 487
   |||||
1340 GCGGGAACGGCTTGTGTGAGCGCAGGGAGGAGTGC.....GACTGCGGC 1383
   |||||
488 SerGlyTYrGlnThrCysLeuAlaPheArgThrPheGlnProCysLysG 504
   |||||
1384 TCGGTGCAGAG..... 1395
   |||||
504 nLeuTrpCysSerHisProAspAsnProTYrPheCysLysThrLysG 521
   |||||
1396 .....TGCAGCCG..... 1404
   |||||
521 lYrProProLeuAspGlyThrGlyCysAlaProGlyLysTrpCysPheLys 537
   |||||
1405 .....GCAAGTGGCAACTGCTGCAAGAA 1428
   |||||
538 GlyHisCysIleTrpLysSerProGlnGlnTrpTYrGlyGlnAspGly 554
   |||||
1429 .....TGC.....ACCTGACTCAGAGAGCCAT 1451
   |||||
554 YTrpSerSerTrpThrLysPheGlySerCysSerArgSerCys..... 568
   |||||
1452 GTGCACAGAC.....GGCTGTGCTGTGGCGGCTGCAAGTACG 1489
   |||||
569 .....GlyGlyGlyValArgSerArgSerArgSerCysAsnAsnPro 582
   |||||
1490 AACCAAGGGGTGTCTCTGCGAGAGGCCGTGAGACAGTGGACATC... 1536

```

583	serProLaTyGtYgLYaTgPrOCysLeuGlyPROMetPheGluTyrGcl	599
1537	.....GCGAAGACCTGCACCGGG.....	1554
599	nValCysAsnSerGtUGtUGtCysPrOGtLYthTyrGluAsPheATaG	616
1555	.....GACTGTACCAGTCCCGCTAACCTGCAC.....	1584
616	InGInCysAlaLysArGAsnSerTyrTyrValInGInAsnAlaLysHis	632
1585	.....AAGTGGACGGTAACTACTCT.....	1605
633	serTTPrValProTyrGluPrOCAsPAsPAlaGInLysCysGluLeuI	649
1606	.....GACCATGACGACGGCGCCGTGCTAC.....	1629
649	eCysGInSerAlaAsPThnGlyAsPArValAlPheMetAsnGlnValAln	666
1629	.....	1629
666	IsAsPglYthArGcYsSerTyrArGAsPPrOTyrSerValCysAlaArG	682
1630	.....GGAGGTGCGTGCACAAACCGGACCGGAC.....	1659
683	GlyGtUGtCysValPrOVaLGtCysAsPbryGInValGtGlySerMetLysAl	699
1660	.....TGCCAGGTGCTTGGGCGCCATGCGGCTGC	1688
699	aAsP.....AsPbryCysGtLYaLcysGtLYaLcysAsnSerH	712
1689	TGATTCGCTCTGCTACAGAACCTAAATGTGGAGGGACGGAGCGTGGA	1738
712	IsCysArGtThrVallyGtGtLYthLeuGtLYsAlaSerLysGln.....	726
1739	GCGTGGCGGCAAGAGATCGGCTGGTCCAGTGCAGTAAACGACGACGCTG	1788
727	..AlaGtLYaLcysLeuValGlnTLeProlaGtAlaArGHisI	742
1789	CTGTGTGGCTCTCTCTCTGTGTCAACATCTGTGAGCTCTCGG.....	1833
742	eGlnInGtAlaLeuGlnLysSerProHisArGSerValYAlLysAsnG	759
1834	.....CTAAGGG	1840
759	InValThrGtLysSerPhe.ILeuAsnProLysGtLYsGlnAlaThrs	775
1841	ACGTGGGGAGACATCACTACTGTGTACCTTGCACCAAGGAGAGAG	1890
775	rArGtThrPheThrAlaMetGtLYeUGtGtTPrGlnAsPArValaGlnAsPAr	792
1891	CTGACAGTCGAGGGGACGACGACGTCGAGCGGCGGCTGTGACGTGAG	1940
792	LaLysGtLysSerLeu.....	796
1941	CTATGTGGAGAGATGCACAGACCTGGCGGCGCTAACATGTGTGTGCTGGAC	1990
797	.....LysThrSerGtLYProLeuProGlnAl	805
1991	ATGCGTCCCTGCCAGCTTCTGCCCTTAACTTGAACCTGCCCC..GC	2037
805	AlaAlaAlaLeuAlaLeuProThrGtGlnGtLYeLYProArGSerSerI	822
2038	AGTGGGAGCGCGGAGTATGTGTCGCCACACGCGGG.....TCGTG	2075
822	euaLArTLYsTYrValIleHisGlnAsPLeuLeuProLeuIleGlySer	838
2076	CACCAATGAAG.....GGAAGT	2092
839	AsnAsnValaLeuGtUGtUGtMetAsPThrTYrGtLYrAlaLeuLysSe	855
2093	GCATCTGTCTC.....AGCCAGACTGGACACGCAAGAACTGCAGTA..TC	2133

[illegible]

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-477-407-8

## seq documentation block:

Sequence 8, Application US/08477/407  
 Patent No. 5631351

## GENERAL INFORMATION:

APPLICANT: NAKAMURA, YUSUKE  
 APPLICANT: EMI, MITSURU  
 TITLE OF INVENTION: MDC PROTEINS AND DNAs  
 TITLE OF INVENTION: ENCODING THE SAME  
 NUMBER OF SEQUENCES: 20  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS P.C.

STREET: 2026 Rambling Road

CITY: Kalamazoo

STATE: Michigan

COUNTRY: USA

ZIP: 49008-1699

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inches, 1.44 MB storage

COMPUTER: IBM PC/XT/AT Compatible

OPERATING SYSTEM: MS-DOS 5.0

SOFTWARE: Wordperfect 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477, 407

FILING DATE: 07-JUN-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/243,542

FILING DATE: 13-MAY-1994

APPLICATION NUMBER: JP 5-136602

FILING DATE: 14 MAY 1993

APPLICATION NUMBER: JP 5-257455

FILING DATE: 22 SEPTEMBER 1993

APPLICATION NUMBER: JP 6-49904

FILING DATE: 23 FEBRUARY 1994

APPLICATION NUMBER: JP 6-73328

FILING DATE: 12 APRIL 1994

APPLICATION NUMBER: JP 6-84470

FILING DATE: 22 APRIL 1994

ATTORNEY/AGENT INFORMATION:

NAME: Terrence F. Chapman

REGISTRATION NUMBER: 32 549

REFERENCE/DOCKET NUMBER: Futuya Case 1313

TELEPHONE: (616) 381-1156

TELEFAX: (616) 381-5465

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 3183 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

LIBRARY: human fetal brain cDNA library

## FEATURE:

NAME/KEY: 3' UTR

LOCATION: 2308..3183

NAME/KEY: CDS

LOCATION: 1..2307

US-08-477-407-8

## alignment\_scores:

Quality: 267.00

Ratio: 0.560

Percent Similarity: 39.883

Percent Identity: 21.321

## alignment\_block:

US-09-938-330-20 x US-08-477-407-8

Align seg 1/1 to: US-08-477-407-8 from: 1 to: 3183

```

59 GLYPROLAlaLaLa.....Serial 65
76 GGCTGCTGAGAGCTTCGATGGGGGGCTTACCCAGTCGGAGGCC 125
65 AGISermetValValAspThrProThrLeuProArgHisSer 81
126 AGAGAGCCCTGAGTGTACGAGAACCCAGCCGTCTGTTAGGAGAGCTCCG 175
82 .....HisLeuArgValAlaArgSerPro 89
176 GGGGAGAGGTCCGAAGCAGCAGCTGACACAGAGGTCCCGCAGACCA 225
90 LeuHisProGlyGlyThrLeuTrpProGlyArgValGlyArgHisSerLe 106
226 .....CCAGGGGGGCCG.....CTGTCCATCTGGCCCGAGTCACT.. 261
106 uTyrPheasnValThrValPheGlyLysGlyLeuHisLeuArgLeuArgp 123
262 ....TTGCTATCCCGACCTCAACTCAACTCACTCACTGACCTGGAGC 307
123 roAsnArgArgLeuValValProGlySerSerValGluTrpGluAsp 139
308 TGAACCAACCACTCTCTC.....TCTGCAATACGTGGAG... 342
140 PheArgGlyLeuPheArgGlyProLeuArgGln..... 150
343 ...GCCCACTTCAGCCGGGAGGAGCAACCCAGCAGCAGCCGGGCTGG 389
151 ...GluCysValTyrThrGlyGlyValThrGlyMetProGlyAlaAlaIav 166
390 AGACCACTGCTACTACAGGAGGAGCTCCGGGAGACCCGACCTCTTGC 439
166 aAlaIleSerAsnGlyAspGlyLeuAlaGlyLeuIleArgThrAspSer 182
440 CGCCCTCTCCACCTGCGGAGGAGCTGCATGGCTCTTGTGATGGAGC 489
183 ThrAspPhePheIleGluProLeuGluArgGlyGlnGlnIuysGluAl 199
490 TTGACTTACATCGTGAAGCCCAAGAG..... 516
199 aSerGlyArgThrHisValValTyrArgArgGluAlaValGlnGlnIu 216
517 .....GTGCTGGACCTT 529
216 rPalGluProAspGlyAspLeuHisAsnGluAlaPheGlyLeuGlyAsp 232
530 GGGGAGCCCTCAG.....GACCC 549
233 LeuProAsnLeuLeu.....GlyLeuValGlyAspGlnLeuGlyAs 246
550 CTTCCCAACCTCATTTACCGGACCCCTCTCCCAAGATCCCTCGAG 599
246 pThrGlu..... 248
600 CAGGGAACAGAGCTGCTGTTGCTGCTGCCAGTGCGCTCTCCAA 649
249 .....ArgLysArgArgHisAlaLysProGlySerTyrSer 260
650 ACCGCGCAGGCTGAGAAAGGAGGAGTCCGCGGGGCCACCTTACA 699
261 IleGlu.....ValLeuLeuValValAspAspSerVa 271
700 GTGCACAGTAACCAAGATGTGAGCTAATTGTGATCAAGACACCA 749
271 lValArgPheHisGlyLysGluHisVal.....GlnAsnTyrValL 285
750 GCTGTGAGCAGATGAGACAGTGGTGTCTCTCAGCAGACATTTGCCA 799
285 eutHrLeuMetAsnIleValAspGluIleTyrHisAspIuSerLeuGly 301

```

```

800 AGTCCGTGTGTAACCTGGCCATGTGATATACAGAGCAG..... 840
302 ValHisIleAsnIleAlaLeuValArgLeuIleMetValGlyTyrArgI 318
841 .....CTCAACACTGCGATCGTCCTGGTTGCCATGGAAACATGGGCACA 884
318 nSerLeuSerLeuIleGluArgGlyAsnProSerArgSerLeuGluGlnV 335
885 TGGGGACAAAGATCCAGGTGACCTCTGGAGACCTGGCCGGCCGGC 934
335 aLysAlaGlyThrAlaHisSerGlnGlnArgGlnAspProSerHisAlaGlu 351
935 TCATGTGTCTAC...CGAGGGAGGGTGTGCTTACGCCCATATGCGCAC 981
352 HisHisAspHisValAlaPheLeuThrArgGln...AspPheGlyProSe 367
982 CAC.....CTTCTTGGCGGACGACCTTCCAGAGCAGAGCAG 1019
367 rGlyTyrAlaProValThrGlyMetCysHisProLeuArgSerCysAlaL 384
1020 CGGGGACACCTACCTGGGGGGCATA.....TGCTCC 1051
384 euAsnHisGluAspGlyPheSerSer.....Ala 393
1052 TGTCCTATGGCGGGGGGTGTGACAGATACGGCAACATGGGGCGGATGCC 1101
394 PheValIleAlaHisGlyThrGlyHisValLeuGlyMet.....G 407
1102 GTGACACCTTGGCCAGACGCTGGAGACAGACCTGGCATGTGACAA 1151
407 uHisAspGlyGlnGlyAsnGlyCys...AlaAspGlyThrSerLeuGlyS 423
1152 ACACCGGACGTCGGGAGGGGACTCCAAAGTCCAGACATCTGCTGGCT 1201
423 eValIleMetAlaProLeuValGlnAlaAlaPheHisArgPheHisTrpSer 439
1202 GCATCATGTGAGGACACCTGGTTTACCTGCCCGCCAAAGTTC.....TCT 1245
440 ArgCysSerIleValLeuGlnLeuSerArgTyrLeuPro.....SerTyrAs 454
1246 CCGTCGACAGATCGACAGATACACCAAGTTCTTCAGAGGGGTGTGCGAG 1295
454 pGlySerLeuAspAspProPheAspProAlaTrpProGlnProProGluL 471
1296 CTGGCCCTTTCACAAACCCCTCAAGCTCTG.....GACCCCCCAAGT 1339
471 euProGlyIleAsnTyrSerMetAspGlnGlnCysArgPheAspPheGly 487
1340 GGGGGACGGCTTGTGGAGGGCAGGGGAGAGTGC.....GACTGCGCG 1383
488 SerGlyTyrGlnThrCysLeuAlaPheArgThrPheGluProCysLysG 504
1384 TCGGTGACAGAG..... 1395
504 nLeuTyrCysSerHisProAspAsnProTyrPheCysLysThrLysLysG 521
1396 .....TGCAGCCGC..... 1404
521 LyrProProLeuAspGlyThrGlnCysAlaProGlyLysTrpCysPheLys 537
1405 .....GCAGGTGGCAACTGCTGCAAGAA 1428
538 GlnHisCysIleThrLysSerProGluGlnThrTyrGlnAspGlyLys 554
1429 .....TGC.....ACCTGACTCAGAGCCCAT 1451
554 yTrpSerSerThrLysPheGlySerCysSerArgSerCys..... 568
1452 GTGCACGAC.....GGCTCTGCTGTGCGCGCTGCAAGTACG 1489
569 .....GlyGlyGlyValArgSerArgSerArgSerCysAsnAsnPro 582
1490 AACCAAGGGGTGTGTCTGTCGAGAGGCCGTGAACGAGTGGCAGATC... 1536
583 SerProAlaTyrGlyGlyArgProCysLeuGlyProMetPheGluTyrG 599
1537 .....GGAGAGCTGTGCACCGGG..... 1554
599 nValCysAsnSerGlnGlyCysProGlyThrTyrGlnAspPheArgAlaG 616
1555 .....GACTTACGCCAGTGGCCCGCTTAACTGCGAC..... 1584
616 lngIncysAlaLysArgAsnSerTyrTyrValHisGlnAsnAlaLysHis 632
1585 .....AAGCTGACGGTTACTACTGT..... 1605
633 SerThrValProTyrGlnProAspAspAlaGlnLysCysGluMetI 649
1606 .....GACCATGACAGAGCGCGCTGCTAC..... 1629
649 eCysGlnSerAlaAspThrGlyAspValAlaPheMetAsnGlnValAln 666
1629 ..... 1629
666 tsAspGlyThrArgCysSerTyrArgAspProTyrSerValCysAlaArg 682
1630 .....GGAGTGGCTGCAAAACCGGGAGCGCAG..... 1659
683 GlyGlnCysValProValGlyCysAspLysGlnValAlGlySerMetLysAl 699
1660 .....TGCCAGGTCTTTGGGGCCATGGCGCTGC 1688
699 aAsp.....AspLysCysGlyValLysGlyGlyAspAsnSerH 712
1689 TGATCGCTTCTGCTACGAGAACCTGATGTGAGGGAGCGAGCGCGGGA 1738
712 tsCysArgThrValLysGlyThrLeuGlyLysAlaSerLysGln..... 726
1739 GCTGTGGCGGCAAGAGATCCGGCTGGGTCCAGTGCAGTAAAGCAGACG 1788
727 ..AlaGlyAlaLeuLysLeuValGlnIleProAlaGlyAlaArgHisI 742
1789 CTGTGTGCTCTTCCTGCTGTGTCAACATCTGTGAGACTCTGCG..... 1833
742 eGlnIleGlnAlaLeuGlnLysSerProHisArgSerValValLysAsnG 759
1834 .....CTAGGG 1840
759 lValThrGlySerPheIleLeuAsnProLysGlyLysGlnAlaThrSe 775
1841 ACCTGGTGGAGACATCAGTGTGTCACTTTCACACGAGGCGAGAG 1890
775 rArgThrPheThrAlaMetGlyLeuGlnTyrGlnAspAlaValGlnAsp 792
1891 CTGACCTCAGGGAGGAGCGACGTCAGCTGGCGGCGCTGTGACCTGAG 1940
792 lalysGlnSerLeu..... 796
1941 CPATGTGAGGATGGCACAGCTCGGAGCTAACATTTGCTGCTGAGAC 1990
797 .....LysThrSerGlyProLeuProGlnAl 805
1991 ATCGCTGCTGCGACCTTCTGCTTCAACTTCACACCTGCGCGG...GC 2037
805 aIleAlaIleAlaLeuAlaLeuProThrThrGlnGlyGlyProArgSerI 822
2038 AGTGGGAGCGCGGATTTGCTCCACACGCGG.....TCTG 2075
822 euAlaTyrLysTyrValIleHisGlnAspLeuProLeuIleGlySer 838
2076 CAGCAATGAG.....GGAAGT 2092
839 AsnAsnValLeuLeuGlnGlnMetAspThrTyrGlnTyrAlaLeuLysSe 855
2093 GCATCTGTC.....AGCCAGACTGTGACAGCAAAAGCTGAGTA...TC 2133

```

```
855 rTTPAlaProCysSerLysAlaCysGlyGlyIleGlnPheThrLysT 872
      |:::|||||
2134 CATACCCCTGCGCAAGTCCACCACGGAGG..... 2167
      |||||
872 YrGlyCysAlaGArgArgAspHisMetVal.GlnArgHisLeuCy 888
      |||||
2168 .....AGACGAGAGAT...ATAAGTCCACGCGCACCAACA 2203
      |||
888 sAspHisLysLysArgProLysProIleArgArgCysAsnGlnHisP 905
      |||
2204 TCATCAT..... 2210
905 rOCysSerGlnProValTrpValThrGluLutrpGly.....AlaCys 919
      ||:::|||||
2211 .....TGAGTC...CATCGTGGGGCTGCTGCTGTC 2240
920 SerArgSerCysGlyLysLeuGlyValGlnThrArgGlyIleGlnCysLe 936
      ||:::|||||
2241 AGCATCTGCTCTGGCGGCACGGGCTG..... 2267
936 uLeuProLeuSerAsnGlyThrHisLysValMetProAlaLysAlaCysA 953
      |||||
2268 .....GGGATTTAAACATTCCGCCGAGAGAGTCCGGA 2301
953 laGlyAspArg.ProGlnuAlaArgArgProCysLeuArgValProCysPr 969
      |||
2302 GGGGCCCTAAGTGCACCCCTGC..... 2322
969 oAlaGlnTrpArgLeuGlyAlaTrpSerGlnCysSerAlaThrCysGlyG 986
      |||||
2323 .....TCCCTCCAAAGCTGGCACCCACCGCT..... 2349
986 lueGlyIleGlnGlnArgGlnValCysArgThrAsnAlaAsnSerLeu 1002
2349 ..... 2349
1003 GlyHisCysGluGlyAspArgProAspThrValGlnValCysSerLeuPr 1019
      |||||
2350 .....CGGCCCTGAACCAAGAG.....CTGCC 2372
1019 o.....AlaCysGlyGlyAsnHisGlnAsnSerThrValArgAlaAspV 1034
      |:::|||||
2373 CCCATCCAGCCACGAGGAGGAGCAACATGATGTCCTCCAGSTCCAAA 2422
1034 aLTrpGluLeuGlyThrProGluGly..... 1042
      |:::|||||
2423 CCTTCAACTCTGCTCGCAGGGGTTGGGTGGGGCTGTGGCCCTGC 2472
1043 GlnTrpValProGlnSerGlyPro..... 1050
      |||
2473 CCTTGGACACCAAGGAGTGGACCAAGCCCTGGAGGGCACTTCTCCACAGT 2522
1051 LeuHisProIleAsnLysIleSerSerMetCysAlaAlaGluProCysT 1067
      |||||
2523 CCCCACCCA.....CCTCGCGGCTCAGCCTTGCA 2554
1067 hngGlyAspArgSerValPheCysGlnMetGlnValLeuAspArgTyrCys 1083
      |||||
2555 CACCCACTGCCCGTGAATGATGCTTCCACCTCATGATGTCACAGC 2604
1084 SerIleProGlyTyrHisArgLeuCysCysValSerCysIleLysLysAl 1100
      |||||
2605 TCAACTCGGGGGCACCTGGAGGATGCCCCAGACAGC.....AC 2645
1100 aSerGlyProAsn.....ProGlyPro...Asp 1109
      |:::|||||
2646 CAGTGGACCTAGCCTGATGAGCCCTCTTGCACACAGGAGGAGCTGAGACC 2695
1109 rOGlyProThrSerLeuProProPheSerThr 1119
      |||||
2696 AGGCTTATCTCTGTGGACTAGGGGAGC 2727
```

